American Psychological Asso
Psychological monographs :
R.W.B. JACKSON LIBRARY
OISE CIR
3 0005 03033 2798

VOL. XLVIII
No. 1
PSYCHOLOGICAL REVIEW PUBLICATIONS
WHOLE NO. 213
1936

Psychological Monographs

EDITED BY
JOHN F. DASHIELL, UNIVERSITY OF NORTH CAROLINA

S. W. FERNBERGER, UNIVERSITY OF PENNSYLVANIA *(J. Exper. Psychol.)*
W. S. HUNTER, CLARK UNIVERSITY *(Psychol. Index)*
H. S. LANGFELD, PRINCETON UNIVERSITY *(Psychol. Rev.)*
JOHN A. McGEOCH, WESLEYAN UNIVERSITY *(Psychol. Bull.)*

STUDIES IN THE PSYCHOLOGY OF ART

Volume II

EDITED BY
NORMAN C. MEIER

HAMILTON SMITH LIBRARY
DISCARDED
FROM
UNH LIBRARY

UNIVERSITY OF IOWA STUDIES IN PSYCHOLOGY

CHRISTIAN A. RUCKMICK, EDITOR

No. XIX

PUBLISHED FOR THE AMERICAN PSYCHOLOGICAL ASSOCIATION BY
PSYCHOLOGICAL REVIEW COMPANY
PRINCETON, N. J.
AND ALBANY, N. Y.

150.8
P974
v.48 no.1

Publications of the American Psychological Association

EDITED BY

S. W. FERNBERGER, UNIVERSITY OF PENNSYLVANIA (*J. of Exp. Psych.*)
WALTER S. HUNTER, CLARK UNIVERSITY (*Index and Abstracts*)
HENRY T. MOORE, SKIDMORE COLLEGE (*J. Abn. and Soc. Psychol.*)
HERBERT S. LANGFELD, PRINCETON UNIVERSITY (*Review*)
JOHN A. McGEOCH, WESLEYAN UNIVERSITY (*Bulletin*)
JOHN F. DASHIELL, UNIVERSITY OF NORTH CAROLINA (*Monographs*)

HERBERT S. LANGFELD, Business Editor

PSYCHOLOGICAL REVIEW

containing original contributions only, appears bi-monthly, January, March, May, July, September, and November, the six numbers comprising a volume of about 540 pages.

PSYCHOLOGICAL BULLETIN

containing critical reviews of books and articles, psychological news and notes, university notices, and announcements, appears monthly (10 numbers), the annual volume comprising about 720 pages. Special issues of the BULLETIN consist of general reviews of recent work in some department of psychology.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

containing original contributions of an experimental character, appears bi-monthly, February, April, June, August, October, and December, the six numbers comprising a volume of about 900 pages.

PSYCHOLOGICAL INDEX

is a compendious bibliography of books, monographs, and articles upon psychological and cognate topics that have appeared during the year. The INDEX is issued annually in June, and may be subscribed for in connection with the periodicals above, or purchased separately.

PSYCHOLOGICAL ABSTRACTS

appears monthly, the twelve numbers and an index supplement making a volume of about 700 pages. The journal is devoted to the publication of non-critical abstracts of the world's literature in psychology and closely related subjects.

PSYCHOLOGICAL MONOGRAPHS

consist of longer researches or treatises or collections of laboratory studies which it is important to publish promptly and as units. The price of single numbers varies according to their size. The MONOGRAPHS appear at irregular intervals and are gathered into volumes of about 500 pages.

JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY

appears quarterly, June, September, December, March, the four numbers comprising a volume of 448 pages. The journal contains original contributions in the field of abnormal and social psychology, reviews, notes and news.

ANNUAL SUBSCRIPTION RATES

Review: $5.50 (Foreign, $5.75). **Index:** $4.00 per volume.
Journal: $7.00 (Foreign, $7.25). **Monographs:** $6.00 per volume (Foreign, $6.30).
Bulletin: $6.00 (Foreign, $6.25). **Abstracts:** $6.00 (Foreign, $6.25).
Abnormal and Social: $5.00 (Foreign, $5.25). Single copies $1.50.
Current numbers: **Journal,** $1.25; **Review,** $1.00; **Abstracts,** 75c; **Bulletin,** 60c.

COMBINATION RATES

Review and Bulletin: $10.00 (Foreign, $10.50).
Review and J. Exp.: $11.00 (Foreign, $11.50).
Bulletin and J. Exp.: $12.00 (Foreign, $12.50).
Review, Bulletin, and J. Exp.: $16.00 (Foreign, $16.75).
Review, Bulletin, J. Exp., and Index: $19.00 (Foreign, $19.75).

Subscriptions, orders, and business communications should be sent to the

PSYCHOLOGICAL REVIEW COMPANY

PRINCETON, N. J.

VOL. XLVIII
No. 1
PSYCHOLOGICAL REVIEW PUBLICATIONS
WHOLE NO. 213
1936

Psychological Monographs

EDITED BY
JOHN F. DASHIELL, UNIVERSITY OF NORTH CAROLINA

S. W. FERNBERGER, UNIVERSITY OF PENNSYLVANIA *(J. Exper. Psychol.)*
W. S. HUNTER, CLARK UNIVERSITY *(Psychol. Index)*
H. S. LANGFELD, PRINCETON UNIVERSITY *(Psychol. Rev.)*
JOHN A. McGEOCH, WESLEYAN UNIVERSITY *(Psychol. Bull.)*

LIBRARY
MAY 11 1973
THE ONTARIO INSTITUTE
FOR STUDIES IN EDUCATION

STUDIES IN THE PSYCHOLOGY OF ART

Volume II

EDITED BY
NORMAN C. MEIER

UNIVERSITY OF IOWA STUDIES IN PSYCHOLOGY

CHRISTIAN A. RUCKMICK, EDITOR
No. XIX

PUBLISHED FOR THE AMERICAN PSYCHOLOGICAL ASSOCIATION BY
PSYCHOLOGICAL REVIEW COMPANY
PRINCETON, N. J.
AND ALBANY, N. Y.

PREFACE

Two years ago the University of Iowa Studies in Psychology included an integrated volume of psychological investigations envisaging the field of plastic and graphic arts. The present volume gives continuity to the progressive researches in that division of research.

The test of any new endeavor in psychology is its productivity on a high qualitative plane. Certainly by this criterion Professor Norman C. Meier, who has directed all of these studies, deserves a great measure of credit. By instilling enthusiasm in his graduate students, by planning and articulating the relatively independent investigations in a program of research extended over the years, and by circumspectly regarding the major problems from a scientific angle in a territory that is full of romanticism and humanistic evaluations, he has established his reputation as a critical psychologist who is also sympathetically versed in the plastic and graphic arts. It is a difficult task to apply qualitative distinctions and quantitative scores to a region that naturally resists the refinements of science. The reader will therefore notice that the attitude of these published researches is not to wipe away æsthetic appreciation and artistic creation, but to understand them better from the scientific angle and to learn their intricacies within the limits of scientific possibility.

The Editor of the Series is sincerely grateful to the several authors for their willingness to prepare their articles in conformity with the established standards of publication in this Series. To Dr. Norman C. Meier the General Editor is especially indebted for preliminary editing of the papers and for many suggestions looking toward the improvement of our Iowa Studies.

Christian A. Ruckmick

Iowa City, Iowa
15 April, 1936

INTRODUCTION

by

NORMAN CHARLES MEIER

The eight studies of this volume are part of a research program known as "Genetic Studies in Artistic Capacity" sponsored by the Carnegie Foundation for the Advancement of Teaching and supported by funds of the Carnegie Corporation of New York. Acknowledgment is made to Dr. Frederick P. Keppel, president, and to the late President Henry Suzzallo of the Foundation and to President Walter A. Jessup, his successor. Without such interest and support these studies could not have been made. The program is under the direction of the writer.

The studies follow those of a previously published volume,[1] some being directly related to earlier studies reported therein. A third volume will contain further studies now in progress.

The research program was projected as a series of investigations, each contributing some advance toward one of the following problems: the nature and variation in the general population of artistic capacity and art ability (productivity); the conditions and limitations of learning in art education during the grade school and high school periods; the nature of the appreciation-process and conditions affecting its functioning; and new approaches to the study of creative mind. The procedure has involved a fruitful merging of psychological, art and educational viewpoints and personnel, utilizing theoretical, experimental, statistical and case study techniques. Theoretical discussion and planning is usually followed by experimental developments. When the latter reach a point where apparent findings are in sight, the procedure leads to application to larger numbers of subjects in the schools. Co-operation has been secured from the public schools of Cedar Rapids, Villisca, Grinnell, West Liberty,

---

[1] Studies in the Psychology of Art, *Psychol. Monog.* 45, 1933, No. 1.

Washington, and Des Moines, Iowa; Lincoln and Cozad, Nebraska; and Milwaukee, Wisconsin. Grateful acknowledgment is made to the superintendents, principals and teachers and particularly to Estelle Hayden, the Director of Art, of Des Moines and to Alfred Pelikan of Milwaukee, since access to subjects such as they provided was an essential condition for the development of the program.

All of the investigations used children as subjects. Consequently all techniques and materials had to be adjusted or adapted to comprehension at the child-level. This the reader should bear constantly in mind.

The first study by *Voss* was projected with a view toward discovering how the child in the grades secured whatever measure of art appreciation he may at that time possess. It is conceivable that appreciation might be simply absorbed from the environment without instruction. It is an educational tenet that contact with pictures or reproductions of famous paintings is instrumental in inculcating appreciation. Furthermore, various conditions affect the process, such as participation either in the active study of pictures, or in producing some kind of art work. The question, furthermore, as to what extent appreciation is possible in the general population is something not at present known. These are some of the questions which the *Voss* investigation attempted to study. Such considerations are obviously basic to any intelligent plan of art education and constitute a neglected aspect of educational psychology.

The *Walton* study of empathic responses in children carries the studies made by *Lundholm, Poffenberger, Guilford, Roy* and others down to the child-level with simplified materials adapted to the child-range of response, in order to discover whether or not empathic response does occur in children and the extent to which it becomes significant at each age-level. While empathic response is a special phase of appreciation, it is believed by a number of psychologists and art educators to be a basic consideration in disclosing artistic sensitivity. There is, undoubtedly, value in the study of the emotional suggestion of line, color, and facial expression as elements in art; and the Walton study shows that it has some differentiating possibilities.

The second *Voss* study is included in this volume as a contribution toward the reliability and validity of a modified form of the *McAdory* test applied to the lower grade levels. Heretofore, the *Meier-Seashore,* the *McAdory,* and other art tests have been considered primarily as instruments for use on adults but of doubtful value with children below the age of ten. In making certain changes in the manner in which materials are presented to the child, *Voss* has adapted the test for use at lower ages.

The study by *Tiebout* on the measurement of quality in children's paintings by the scale method describes the procedure followed in the construction of a scale for experimental purposes in connection with a number of current studies. The necessity of having some objective basis for evaluating children's productive ability and to provide a set of measurements that might then be correlated with intelligence, æsthetic sensitivity, progress in art ability from year to year, and other considerations, led to the construction of this scale through the co-operation of Des Moines schools and the co-operation also of 14 art educators in various parts of the country who served as judges.

The study of the relationship of artistic ability and general intelligence was approached through three groups of subjects. The productive ability of approximately 100 subjects in each of the seven grades of a typically ' average ' Des Moines school was ascertained by means of the scale and then correlated with general intelligence. The second group of subjects were selected students from the Milwaukee schools that were designated by Director Pelikan, with the assistance of supervisors and teachers, as outstanding in art activities. The third group consisted entirely of prominent American artists selected from a list of 5,500 on the basis of general reputation, honors won, and general character of achievement. The interesting finding of this study is that while results are variable from grade to grade in the grade schools, the average of intelligence for the high school groups is definitely above average, while the average intelligence of the mature artists was found to be 119 at the dividing point between superior and very superior intelligence.

The *Saunders* study was an attempt to determine experimentally whether artistic ability is more or less fixed at birth or early in life; or, if, on the other hand, it is possible to change it by appropriate training. This touches on two widespread beliefs, namely, on the one hand that art ability is rare and more or less unpredictable and hence can only be known by its emergence. The other belief holds that any one can be taught to draw. To make the problem decisive, Saunders undertook to use as subjects a small group of grade school children who had never shown any interest or aptitude in drawing or painting. That he was able to show results which are indisputable and contrary to the expectations of some viewpoints, is remarkable.

In contrast to this experiment is the case study of Loran Lockhart, a boy practically blind from birth, for whom a fair degree of vision was made possible at the age of seven by cataract operations. While cataract cases in young children are not in themselves rare, a case of operative recovery accompanied by definite drawing ability, even under such a handicap as unsatisfactory vision, was believed to be sufficiently rare to warrant investigation. This case was one in which the factor of visible environment approached zero. Unlike the preceding studies the conclusion of this study points to a theory rather than to experimental findings.

The final article is a description of a new approach to the study of creative artistic intelligence. It describes an apparatus or device for fabricating æsthetic organizations and controlling these at will. It is believed,—since almost any condition of light, mass, color, form, and organization can be had by this means, making it possible to duplicate colorings, moods or emotional quality,—that it would provide a flexible device of unlimited possibilities for creative expression and for the construction of experimental compositions for various purposes. This device is now being used by William McCloy in an investigation of the creative imagination of children and will be used in other studies for various purposes. These studies will be described at a later date.

The studies by *Walton* and *Tiebout* are postdoctorate investigations, the ones by *Voss* are both done as separate parts of her doctorate study as also was the investigation of *Saunders*. The construction of the scale involved the participation of various research assistants and most of the ideas in all studies were benefited by criticisms and constructive consideration offered by the Psychology of Art Seminar group.

The writer wishes to acknowledge further a contribution to the cost of publishing the volume made by the Conference for Research in Art of the American Federation of Arts and the Carnegie Corporation and also some partial aid for two of the studies by the Spelman Fund through co-operative arrangements with the Iowa Child Welfare Research Station.

Acknowledgment is also made to Professor Christian A. Ruckmick who, as general editor of the Iowa Studies, rendered service in the preparation of the volume for publication.

CONTENTS

Preface .... Christian A. Ruckmick iii

Introduction .... Norman C. Meier v

A Study of Conditions Affecting the Functioning of the Art Appreciation Process at the Child-Level .... Mildred Dow Voss 1

Empathic Responses in Children .... William E. Walton 40

The Validity and Reliability of a Modified Form of the McAdory Art Test for Use at Lower Grade Levels .... Mildred Dow Voss 68

The Measurement of Quality in Children's Paintings by the Scale Method .... Carolyn Tiebout 85

Artistic Ability and General Intelligence ... Carolyn Tiebout and Norman C. Meier 95

The Stability of Artistic Aptitude at the Child Level .... Aulus W. Saunders 126

Art Ability without Instruction or Environmental Background: Case Study of Loran Lockhart .... Norman C. Meier 155

An Instrument for the Study of Creative Artistic Intelligence .... Norman C. Meier and William McCloy 164

A STUDY OF CONDITIONS AFFECTING THE FUNCTIONING OF THE ART APPRECIATION PROCESS AT THE CHILD-LEVEL[1]

by

MILDRED DOW VOSS

I. Introduction. Two important forms of appreciative ability in graphic art are the ability of the individual to analyze the manner in which a composition has been æsthetically organized, and the ability to recognize the effect of variations in the æsthetic organization of a picture upon its relative merit. The former type of appreciative ability will be designated as æsthetic analysis, the latter as æsthetic judgment.

This investigation concerns itself with the effect of two factors, direction and practice, upon the development of æsthetic analysis and judgment in elementary school children, and upon the development of ability to produce an æsthetically meritorious composition.[2]

It would seem logical to suppose that æsthetic analysis and judgment can best be developed through cultivation of a knowledge of the criteria upon which æsthetic merit is based. These criteria are the presence of relationships between the various art elements (such as form and color), which are described by such terms as balance, rhythm, sequence, and harmony. All of these relationships can be subsumed under the larger term 'principles of art'. A picture possesses æsthetic merit when each element is related to itself and to the other elements in such a manner that the composition would be less satisfying if the existing relationships were added to, subtracted from, or altered in organization (unity).

---

[1] An investigation in the research program, Genetic Studies in Artistic Capacity, directed by Dr. Norman C. Meier and sponsored by the Carnegie Foundation for the Advancement of Teaching with financial aid from the Carnegie Corporation. This study received aid also from the Spelman Fund through co-operative arrangements with the Iowa Child Welfare Research Station.

[2] The term 'direction' is adopted from Maier and Reninger who define it in the following manner: "What the difficulty is seen to be will determine the way the problem is attacked. *This way of starting out on a problem is termed direction.*" (**9**, p. 30.) A picture may be considered a substitute-object or an æsthetic organization. Upon the adoption of one or the other of these viewpoints depends the entire subsequent character of the analytical or evaluative response to a picture. In this investigation the term 'practice' is used in the sense of continued experience over a period of time in dealing with problems of the type that are herein described.

There are certain ways of looking at a picture which are not based upon these criteria. One may look at a picture to determine what is portrayed (subject-recognition); how well the portrayal resembles the actual model (photographic representation); to what extent the object interests the observer (subject-interest); what message is conveyed by the situation in which the object is portrayed (literary interest); or how well the setting of the object conforms to those conditions usually accompanying its occurrence (historical or naturalistic accuracy). All of these ways of looking at a picture are characterized by a common approach. The picture is perceived as a representation of an object which is henceforth considered in the same manner as the actual object would be considered were it present—that is, the picture becomes a *substitute-object.* This attitude is opposed to the approach which considers the object portrayed as more or less incidental and regards as important the degree to which the picture is a *unified organization* of pictorial elements. From this character of opposition it would seem to follow that these two ways of looking at a picture are mutually exclusive. As long as the picture is understood as a *substitute-object,* the æsthetic organization is disregarded. When a picture is understood as an *æsthetic organization,* the subject-matter is secondary.

Accordingly one of the principal obstacles in the development of æsthetic analysis and judgment occurs at the beginning of the process. A picture must be perceived in terms of an organization of æsthetic elements in order that development may occur. Attitudes which inhibit such an understanding of the problem must be broken down. Specifically, this investigation proposes to test the hypothesis that an important step in the development of æsthetic analysis and judgment consists of the revision of attitudes in which good art is perceived as portrayal of interesting subject-matter, photographic representation, evidence of technical ability, a peg upon which to hang associated content, or a story-telling device; and the substitution of the attitude that good art involves selection, elimination, and arrangement according to principles such as unity, harmony, and rhythm.

Once the comprehension of art as an organization of pictorial elements is established, the factor of practice, or experience in the perception of æsthetically significant relationships, becomes important. As the principles of art are generalized relationships which may be exhibited in many concrete forms, adequate understanding requires abstraction of a common relationship from many examples. With repetition of similar types of problems, relationships which are at first perceived with difficulty are recognized more easily upon subsequent reappearance. With mastery of this relationship the attention may be turned to relationships which are as yet difficult to perceive. Accordingly there is an increase in the number, and probably in the quality, of the relationships which can be identified. Eventually more or less complete analysis requires that each element be related through a system of increasingly inclusive relationships to the whole of which it is a part.

With the attainment of at least a partial understanding of the principles, the basis for the development of æsthetic judgment has been established. Upon the presentation of two or more methods of creating an æsthetic organization, each method must be scrutinized as to its degree of success in the attainment of unity. The respects in which the organizations vary and the influence of each variation upon the total effect must be noted, and judgment made on the basis of the appropriateness of the variation to the total form of the picture. With increased experience, variations which are at first with difficulty recognized as enhancing or destroying the unity of the composition are, upon repetition of similar problems, recognized with greater ease so that the attention may be turned to problems not yet solved.

With increase in ability to evaluate the unifying qualities of more types of variations there occurs greater accuracy of judgment. The second hypothesis of this investigation is that practice in the recognition of the principles of art is an important factor in the development of æsthetic judgment.

II. The effect of adequate direction upon the development of æsthetic judgment. Experiment I. Experiments I and II attempted to isolate the effect upon æsthetic judgment of the manner in which the problem of looking at pictures was approached. Experiment I undertook to test the assumption that practice in the judgment of differing degrees of merit cannot bring about a significant increase in the development of æsthetic judgment without an initial comprehension of the picture as an organization of pictorial elements.

The subjects of Experiment I were 101 children in the second, third, and fourth grades of the public schools of Villisca, Iowa.[3] The children were divided into two matched groups on the basis of intelligence (IQ), mental age (MA), chronological age (CA), grade, sex, and æsthetic judgment score. IQ and MA were determined by the *Kuhlmann-Anderson* Group Intelligence Test, 4th edition. Æsthetic judgment was measured by a modified form of the *McAdory* Art Test in which judgment was required only between the best and worst degrees of merit of each exercise.[4] Table I presents the mean scores and standard deviations for each of these measures for each group.

The experimental group, Group A, was assembled as a class for 30 min. a day, three days a week for ten weeks. These ten weeks were distributed over a period of twenty weeks through which two weeks of classes alternated with two weeks of rest. Approximately 140 paired or multiple-choice illustrations were presented in which two or more alternative organizations were shown simultaneously. These alternative organizations were similar in every respect except that the arrangement of objects

---

[3] Acknowledgment is made to Mr. H. E. Dow, Superintendent of Villisca Public Schools, and to the principal and teachers co-operating in this study.

[4] For a detailed discussion of the administrative procedure, validity, and reliability of this form of the test, refer to the article in this volume entitled *The Validity and Reliability of a Modified Form of the McAdory Art Test for Use at Lower Grade Levels.*

Table I. *Description of Graups A and B in Experiments I and II*[5]

		No. of Subjects			Art Score		IQ		MA		CA	
Grade	Group	T	F	M	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
II	A	14	9	5	33.9	7.8	102.6	8.1	91.3	6.2	89.2	6.6
	B	13	9	4	33.6	5.7	105.5	8.7	93.8	5.7	89.2	7.0
III	A	21	10	11	35.5	6.4	98.5	17.1	102.1	12.1	104.8	12.1
	B	20	10	10	35.6	7.2	102.0	14.5	102.5	12.4	101.3	6.5
IV	A	16	11	5	36.4	8.6	100.3	10.8	115.8	10.5	115.5	6.5
	B	17	11	6	37.8	6.6	101.5	12.1	115.4	16.4	114.4	7.3

or the form of the objects was changed so that in one member of the pair a principle was exemplified while in the other member the principle was violated. The general form of a number of these illustrations was adapted by the experimenter from illustrations in magazines, newspapers, and children's books.[6] The remainder of the exercises used were selected from the *Whitford* Art Appreciation materials[7], from a preliminary version of the *Christensen* Art Test[8], and from the *Jacobson* Costume Design Booklet.[9] In the experimental period each child was presented with a mimeographed copy of the exercise. The experimenter pointed out the respect in which the exercises differed, taking care, however, not to suggest the æsthetic principle involved. Each child then checked the picture which he would like to have returned to him at the conclusion of the experiment. Plate I illustrates the form in which the exercises were presented.

At the conclusion of the experiment both experimental and control groups were retested by a repetition of the *McAdory* Art Test. To determine if either group had gained significantly in mean æsthetic judgment score during the experiment, the following formula was applied to determine the probable error of the difference between the original and final mean test score for each group.

---

[5] The IQ, MA, and CA scores are in terms of months.

[6] A copy of these exercises is on file at the Psychology of Art division at the University of Iowa.

[7] Whitford, W. G. An introduction to art education. N. Y., Appleton, 1929, plates I, II, III, and IV, exercises 2, 4, 5, 9, and 11.

[8] Christensen, E. C. A test in art appreciation (revision of March, 1927), exercises 3, 4, 6, 8, 10, 11, 12, 13, 16, 18, 20, 21, 23, 24, 25, 26, 28, 36, 38, 39, 40, 41, 46, 47, 48, 49, 52, 53, 54, 56, 58, 59, 60, 62, 65, 66, 67, 70, 71, 75, 76, 77, 83, 89, 90, 91, 92, 93, 98, 99, 100 (Photographic print form).

[9] A copy of this booklet is obtainable at the Psychology of Art division at the University of Iowa. It is described in the article by Jacobson, W. E., An experimental investigation of the basic æsthetic factors in costume design, *Psychol. Monog.*, 45, 1933, No. 1, 147–184.

(1) $P.E._{diff.} = \sqrt{P.E._{x_1}^2 + P.E._{x_2}^2 - 2\,r\,P.E._{x_1}\,P.E._{x_2}}$ (2, p. 286)

$P.E._{x1}$ and $P.E._{x2}$ represent respectively the probable error of the mean scores in the initial and final test for a particular group.

r represents the Pearson product-moment coefficient of correlation between the initial and final score on the *McAdory* Art Test for each subject.

The significance of the difference between the original and final test scores was determined by the formula (2) $\frac{D}{P.E._{diff.}}$ (2, p. 136) in which D represents the obtained difference between the original and final mean test scores and the $P.E._{diff.}$ was obtained as indicated in formula (1). The chances of a true difference greater than zero was obtained by reference to a table of probabilities (2, Table XV).

To determine if the difference between the means of the experimental and control groups on the final test was significant the following formulae were applied. As the effect of the methods compared resulted in different values of the correlation coefficient between initial and final measures for the two groups, the following formula devised by *Wilks* was applied independently to the mean of each sample and the resulting values substituted for the probable errors of the means in the usual formula for the probable error of the difference.

(3) $P.E._{M} = \frac{.6745\,\sigma\sqrt{1-r^2}}{\sqrt{N}}$

$P.E._{M1}$ represents the probable error of the means of matched groups. *r* is the *Pearson* product-moment coefficient of correlation between the measures used as the basis for matching (the initial *McAdory* Art Test score) and the measures for which the mean is computed (the final *McAdory* Art Test score).

(4) $P.E._{diff.} = \sqrt{P.E._{av._1}^2 + P.E._{av._2}^2}$

The results for Experiment I are to be found under that heading in Table II.

These results seem somewhat contradictory. Simple repetition of the test, which can be construed as practice in judgment of varying degrees of merit, did bring about a significant increase in æsthetic judgment score from the initial to the final test in each group, while a considerable amount of practice in additional exercises failed to result in a significant difference between the final scores of the groups. A possible explanation of these results might be the greater familiarity on the part of the child with the experimental set-up. In the first test the children were for the first time exposed to an individual test situation. In the final test, experimental conditions, experimenter, and test exercises were familiar to the child. An element of competition between groups was also introduced as a motivating factor in the final test. It seems probable that the difference between final test scores of the two groups is a more valid indication of the effect of practice upon an increase in æsthetic judgment score than the increment between initial and final test scores of a single group. This difference in final test scores was not significant, a result which would seem to indicate that practice in judging between varying degrees of æsthetic merit without an understanding of the proper criteria to be employed, produced through instruction in the principles of art, will not produce a significant increase in æsthetic judgment.

Experiment II. To determine if the same experimental exercises accompanied by a verbal explanation of the principles of art would produce a significant increase in æsthetic judgment, Experiment II was performed. The same subjects were used in this experiment, but with the groups reversed. The new experimental group met for five days each week, for three consecutive weeks. The class period was 30 min. in length as before. The procedure, except for an abbreviated series of exercises (N=100), was the same as in Experiment I with the variation, that, after the child had indicated his preference and the slips had been collected, the experimenter designated the correct illustration and explained, in terms suitable to the age of the children, the principle of art involved. A typical explanation is presented accompanying Figure 1.

TABLE II. *Significance of the difference between initial and final scores of experimental and control groups on the McAdory Art Test and between final scores of experimental and control groups for Experiments I, II, III, and IV*

Experiment	Group	No. of subjects	First Test			Second Test			Between First and Second Test			Between Second Tests of Each Group		
			Mean	S.D.	P.E.$_{av.}$	Mean	S.D.	P.E.$_{av.}$	r	P.E.$_{diff.}$	Critical Ratio	P.E.$_{diff.}$	Critical Ratio	P
I	E	51	35.22	7.65	.72	38.65	8.49	.80	.64	.65	5.27	1.06	.92	73
	C	50	35.46	7.17	.69	37.68	8.46	.81	.77	.52	4.27			
II	E	38	38.53	7.80	.85	43.34	7.98	.87	.87	.44	11.03	1.00	4.81	99.7
	C	41	38.56	9.00	.95	39.81	8.82	.93	.79	.61	2.06			
III	E	44	101.18	22.40	2.28	116.93	27.44	2.79	.72	1.95	8.08	1.84	4.85	99.7
	C	41	99.37	20.90	2.12	108.00	28.14	2.97	.87	1.53	5.64			
IV	E	40	107.98	28.49	3.04	116.55	24.50	2.61	.74	2.08	4.12	2.37	1.81	89
	C	42	118.00	26.74	2.78	120.83	29.05	3.02	.86	1.48	1.91			

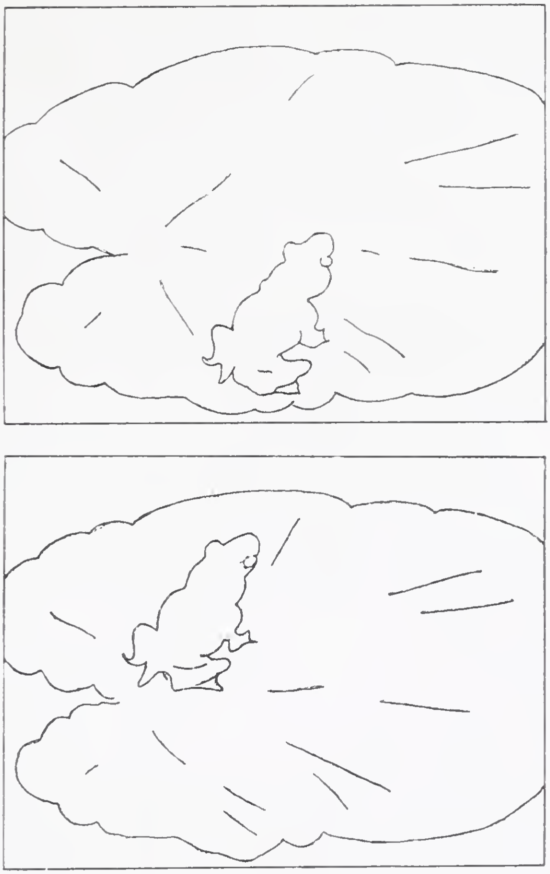


FIG. 1.

FIG. 1. Illustration of the type of experimental exercises used in Experiments I and II

DOMINANCE BY PLACEMENT AT POINT OF RADIATION

Directions for Experiments I and II. In this picture a frog is sitting on a big lily leaf. He is sitting in a different place in each picture. Put the letter of the picture that you like best by number 16 on your paper. Pass your paper to the person back of you.

Further explanation for Experiment II. You remember that we said there must always be a captain of a picture just as there must always be a captain of a football or baseball team. The most important thing, or the thing that we look at most, in a picture is the captain. In this picture we look at the frog most so that he must be the captain. But you remember that when lines came to a point they made a captain too. Now in this lily pad there are lines that lead to a point. In picture A we have one captain here where the lines meet and another captain here at the frog. What happens if we have a team with two captains? It doesn't get along very well, because one captain often wants to do one thing while the other captain wants to do something else. Until one captain gives in, the team can't play ball.

Suppose we make Jack captain of the baseball team of this school. The teams from all of the schools meet and make a town-team. Jack is elected captain of the town-team so that he is captain of the school-team and captain of the town-team both. He is a more important captain now than if he were just captain of the school-team. In picture B, the frog is placed where the lines come to a point. He has been elected captain of the lines and the whole picture both so that he becomes a more important captain. The picture is better just as a team is better the more important the captain is. B is the better picture because it has one important captain rather than two captains that don't get along together.

At the conclusion of this experiment both groups were retested and the results treated by the same statistical methods as in Experiment I. These results, presented in Table II, indicate that there was a significant increase in score on the *McAdory* Art Test in the experimental group but that the increase in test score for the control group was not significant. Furthermore, the difference in final score of experimental and control groups was significantly in favor of the experimental group. From these results the conclusion seems to be justified that it is necessary that the child comprehend the picture as an organization involving those relationships which have been designated as principles of art before experience in judging between varying degrees of æsthetic merit will produce a significant increase in æsthetic judgment.

III. The effect of number of exercises judged upon gain in æsthetic judgment score. During the course of Experiment II illness necessitated the absence of a number of children during portions of the experiment. Accordingly the number of exercises judged by each child varied considerably. The procedure in this experiment consisted of the presentation of several exercises at each class period, each exercise differing in the nature of the principle exemplified. During the course of one lesson, therefore, a number of principles were explained and succeeding instruction consisted, for the most part, in the application of a relatively few principles to pictures of varying content. Thus, a child might attend the class only a few days and acquire a knowledge of the general attitude and fundamental principles being employed. Constant attendance at the class, however, resulted in a much greater amount of practice in the application of this core of ideas to a great variety of examples. Advantage was taken of the large variation in number of days' attendance to determine if the number of exercises checked and explained bore any relation to the amount of gain in æsthetic judgment score.

On the basis of the general shape of the distribution curve of exercises checked, the children were divided into two groups, those having checked more than 90 exercises and those having checked fewer than 90 exercises. The average amount of gain on the *McAdory* Art Test for each group was then determined. The results are shown in Table III.

TABLE III. *A comparison of the group checking more than 90 exercises and the group checking fewer than 90 exercises in amount of gain in Experiment II*

Group	No. of subjects	Mean number of exercises checked	Mean initial McAdory score	Mean IQ	Mean MA	Mean CA	Mean gain
90–100	14	95.7	35.5	102.5	106.9	105.0	6.09
30–90	11	53.5	35.3	97.5	97.5	95.1	2.36

The gain in the group checking over 90 exercises was 258 per cent that of the group checking fewer than 90 exercises. However, it is evident that the IQ, MA, and CA scores were considerably in favor of the group making the greater gain. In order to determine if these factors were affecting this gain in score, the children in the two groups were paired according to each measure and the difference in gain recomputed. For example, a child having an IQ of 104 in the 90 to 100 exercise-group was matched to a child having an IQ of 104 in the 30 to 90 exercise-group. Children were paired only if the measures in question were within one point of each other so that the average and standard deviations of each group were identical. The two new groups thus obtained were then compared for gain in æsthetic judgment score. The same procedure was followed with MA, IQ, and initial *McAdory* Art Test score. These results are presented in Table IV.

TABLE IV. *A comparison of the group checking more than 90 exercises and the group checking fewer than 90 exercises in amount of gain in Experiment II when the groups were paired according to MA, CA, IQ, and initial McAdory Art Test score*

Groups matched for	90–100 exercise-group	30–90 exercise-group	Difference in gain
MA	7.4	3.7	3.7
CA	5.9	1.4	4.5
IQ	6.4	2.8	3.4
McAdory Art Test	6.4	4.4	2.0

Because of the small number of cases it was not feasible to determine the significance of the difference in gains. The fact, however, that the gain was substantially in favor of the group checking the greater number of exercises no matter in which factor the groups were matched seems to justify the conclusion that the difference in gain in æsthetic judgment score was a function of the number of exercises judged.

IV. The effect of similarity of experimental exercises to test items upon gain in æsthetic judgment score. The conclusions from the preceding experiments regarding the functions of knowledge of the principles of art and practice in their recognition might be challenged on the grounds that the experimental exercises were so similar to the test items that the increase in æsthetic judgment score might be attributable to a type of indirect coaching upon the test.

A variation in Experiments I and II afforded a chance to test the specificity of gain on the test items to the type of exercises used in the experimental series. The *McAdory* Art Test is divided into three sub-tests, Form, Notan, and Color, for each of which a separate score may be derived. Form includes exercises dealing with the shape of objects; Notan, with the distribution of light and dark; and Color, with the harmonies and accents of color combination. In Experiments I and II, in addition to the *McAdory* Art Test, two supplementary tests were administered: (1) the *Whorley* Unity Scale of Landscape

Arrangements, the following scale values being presented as paired comparisons, values one and seven, two and six, three and five, for each of the two forms of the Bird-bath and Entrance Scales (17); and the *Lewerenz* Test of Fundamental Abilities of Visual Art, Part I, Test 1, Recognition of Proportion (7).[10]

Only three exercises in the training series of Experiments I and II involved notan or color, while a considerable number of exercises dealt with shape, proportion, and the organization of a picture for the production of unity. The principles of art, however, apply to all of the elements enumerated. To determine if a gain had occurred only on those test items resembling the exercises, the difference between the gains in score of the experimental and control groups was determined for each of these sub-tests. The results are presented in Table V.

Inspection of Table V reveals that, in Experiment I, although the difference in gain on the sub-test, Form, was greatest, the difference in gain on the sub-test, Notan, was only slightly less. Both of these differences approached significance. In Experiment II although the greatest difference in gain occurred on the sub-test, Form, the difference in gain on the sub-test, Color, as well as on the sub-test, Proportion, approached significance. Thus a gain was demonstrated on two types of test items differing in the art element involved from the exercises used in the experimental series. The difference in gains on the sub-test, Unity, was not significant in either experiment.[11]

When Experiments III and IV were devised, certain modifications in the types of exercises were made to determine the justification of the objection that the gain in test score was due to the similarity of items to experimental exercises.

The exercises used in Experiment III consisted of examples illustrating only the application of the principle. The child was

---

[10] As the *McAdory* Art Test alone was used as a measure of æsthetic judgment in Experiments III and IV, the scores from these tests were not included in the æsthetic judgment scores in Experiments I and II. The results are presented for these tests separately in Table V.

[11] As the mean initial scores on this test were high, to gain significantly, the child would be required to judge correctly between the items ranked third and fifth on the scale. The brief experimental procedures were probably not sufficient to bring about such an increment in ability.

TABLE V. *The significance of the difference between the gains of the experimental and control groups on the sub-tests, Form, Notan, Color, Unity, and Proportion, in Experiments I, II, III, and IV*

Experiment		Group E				Group C					Critical	
	No.	Mean	S.D.	P.E.$_{av}$	No.	Mean	S.D.	P.E.$_{av}$	Difference	P.E.$_{diff}$	ratio	P
I F.	51	1.54	3.80	.36	48	.66	1.04	.10	.88	.374	2.35	94.5
I N.	51	1.34	1.02	.10	48	.62	2.80	.27	.72	.288	2.50	95.0
I C.	51	—.26	3.28	.31	48	.58	3.44	.34	—.84	.461	1.82	89.0
I U.	50	1.02	1.94	.19	46	1.02	1.76	.18	.00	.261	.00	50.0
I P.	49	.07	2.32	.22	46	—.02	2.22	.22	.09	.310	.29	58.0
II F.	38	2.61	2.65	.29	43	.26	2.43	.25	2.35	.384	6.12	99.7
II N.	38	.55	2.44	.27	43	.47	2.19	.23	.08	.355	.23	56.0
II C.	38	1.74	2.30	.25	43	.88	2.86	.29	.86	.383	2.23	93.5
II U.	34	.53	1.54	.18	38	.45	1.90	.21	.08	.276	.29	58.0
II P.	33	.91	1.96	.22	40	.33	1.77	.19	.58	.290	2.00	91.0
III F.	42	8.28	12.12	1.26	39	2.76	8.94	.96	5.52	1.584	3.48	99.0
III N.	42	3.72	7.32	.72	39	1.92	7.20	.78	1.80	1.063	1.69	87.0
III C.	42	5.49	7.98	.83	39	3.78	7.53	.81	1.71	1.162	1.47	84.0
IV F.	38	5.67	10.47	1.15	40	4.14	10.02	1.07	1.53	1.568	.98	75.0
IV N.	38	1.11	9.90	1.08	40	—.30	8.97	.96	1.41	1.445	.98	75.0
IV C.	38	3.48	8.04	.88	40	.45	7.20	.77	3.03	1.166	2.60	96.0

required to induce the principle from its application in many specific examples. These illustrations were drawn from a large variety of newspaper and magazine sources and exhibited the principle in widely varying types of subject-matter. It seemed that these training exercises were sufficiently dissimilar to the test exercises that any change in *McAdory* Art Score which might result from the experimental procedure should not be attributable to direct or indirect coaching upon the test exercises.

The children used in Experiments III and IV were fifth-grade children of the public schools of Grinnell, Iowa, a town of about 5,000 population. The children were matched according to grade, sex, IQ, MA, CA, and *McAdory* Art Test score as in Experiments I and II. The standardized four-item form of the *McAdory* Art Test was used instead of the two-item form.

Table VI. *Description of Groups A and B, Grade V, in Experiments III and IV*

	No. of Subjects			Art Score		IQ		MA		DA	
Group	T	F	M	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
A	44	23	21	100.8	22.2	103.9	11.7	131.7	11.3	128.8	8.6
B	41	23	18	99.3	20.3	102.0	11.4	130.0	9.4	128.5	8.4

Table III presents the mean scores and standard deviations for each of these measures for each group.

The experimental group, Group A, met for thirty minutes daily for five days a week for three weeks as a part of the regular grade curriculum. Group B received no training. In this experiment a systematic outline of some of the major principles of art was developed. The children were first presented with a verbal explanation of the principle. After this explanation, a number of the illustrations of the type described above were presented, showing the application of the principle in widely varying types of situations. The experimenter pointed out, in each case, how the illustration exemplified the principle. After a number of principles had thus been presented in the training series a group of illustrations exemplifying one principle was presented and the children were required to abstract the principle possessed in common by all of the illustrations. The outline of the principles which was repeatedly placed before

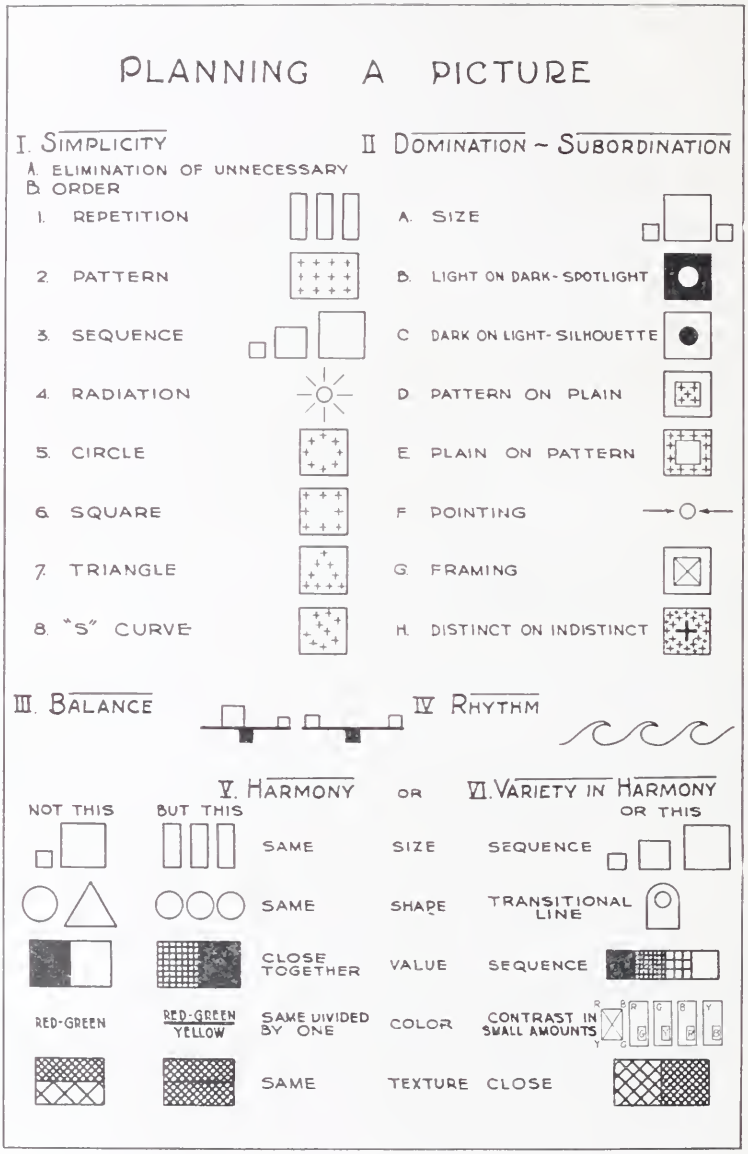

PLANNING A PICTURE
I. SIMPLICITY
A. ELIMINATION OF UNNECESSARY
B. ORDER
1. REPETITION
2. PATTERN
3. SEQUENCE
4. RADIATION
5. CIRCLE
6. SQUARE
7. TRIANGLE
8. "S" CURVE
II DOMINATION ~ SUBORDINATION
A. SIZE
B. LIGHT ON DARK - SPOTLIGHT
C. DARK ON LIGHT - SILHOUETTE
D. PATTERN ON PLAIN
E. PLAIN ON PATTERN
F. POINTING
G. FRAMING
H. DISTINCT ON INDISTINCT
III. BALANCE
IV RHYTHM
V. HARMONY
OR
VI. VARIETY IN HARMONY
NOT THIS
BUT THIS
OR THIS
SAME
SIZE
SEQUENCE
SAME
SHAPE
TRANSITIONAL LINE
CLOSE TOGETHER
VALUE
SEQUENCE
RED-GREEN
RED-GREEN YELLOW
SAME DIVIDED BY ONE
COLOR
CONTRAST IN SMALL AMOUNTS
R
B
G
Y
R
G
G
Y
B
R
Y
B
SAME
TEXTURE
CLOSE

PLATE I.

the children as a basis for discussion is reproduced in Plate I. Plate II illustrates the type of exercises used, and the manner in which these exercises were combined to exemplify a common principle.

Following the experimental period, both groups were retested and the results treated statistically as in Experiments I and II. The results, presented in Table II, demonstrated that training in ability to recognize æsthetic principles in exercises of this type did increase æsthetic judgment significantly.

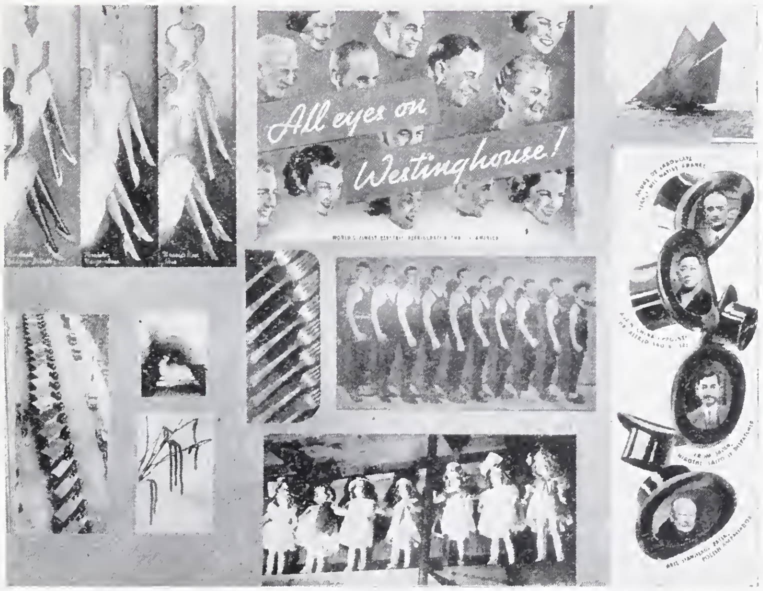

All eyes on Westinghouse!

Plate II.

Experiment IV. Experiment IV utilized exercises similar in type, although different in subject-matter and degree of complexity, to those used in Experiments I and II. As in the former experiments, the child was required to find which of two examples exemplified, which violated the principle, in order to determine which exercise was æsthetically better. Experiment IV can be considered as a repetition of Experiment II under more refined conditions. The experiment was conducted in the fifth grade for which the *McAdory* Art Test is standardized; the exercises were drawn from sources which would insure a more adequate validity; and the explanations of the principles of art were presented in more systematized form.

Group B became the experimental group in this experiment, meeting for the 30 min. each day for five days a week for three weeks as had Group A previously. The children were presented with the same general explanations of the principle under consideration as in Experiment III. Group B, however, was presented with the exercises described above which approach much more closely in general character to the test items than the exercises used in Experiment III. A few illustrations were drawn from magazines and newspapers but the majority of exercises were taken from *Art in Everyday Life* (3) by *Harriett* and *Vetta Goldstein* and from the *Meier-Seashore Art Judgment Test* (12).[12]

Table VII. *The significance of the difference between gains on the McAdory Art Test of Groups A and B in Experiments III and IV*

	Group A				Group B						
Experiment	No.	Mean gain	S.D. gain	P.E.$_{av.}$	No.	Mean gain	S.D. gain	P.E.$_{av}$	P.E.$_{diff.}$	Critical ratio	P
III	44	16.7	18.5	1.88	41	8.9	22.9	2.41	3.06	2.55	95
IV	44	2.0	16.4	1.67	40	8.5	19.5	2.08	2.65	2.45	95

The results, treated statistically as in the previous experiments are presented in Table II. These results indicate that there was a significant increase in mean æsthetic judgment score in the experimental group but not in the control group. The difference between final mean scores of the two groups was not significant. However, the initial mean score of the experimental group was much lower than that of the control group. Accordingly, the difference between the mean gains in score made by the experimental and control groups on Experiments III and IV was computed. The results are presented in Table VII.

---

[12] The general discussion of the principles as presented in Experiments III and IV, and references to the source of each exercise used in Experiment IV are on file at the Psychology of Art division, University of Iowa.

These results indicate that, although both groups made a much larger gain in score on Experiment III, the difference in gain between the two groups and the significance of the difference is the same in both experiments.

It appears, therefore, that although the type of exercises used in Experiments III and IV differed considerably, and although the exercises used in Experiments II and IV were different in subject-matter and degree of complexity, a significant gain was demonstrated in æsthetic judgment score in each experiment.

In Experiment III the exercises involved all of the elements tested in the sub-tests of the *McAdory* Art Test. In general, the amount of emphasis placed on exercises illustrating each element was in the order; form, color, and notan. Notan involved but few exercises. The results, presented in Table V, indicate that the difference in gains between the experimental groups was in the order; form, notan, and color.

In Experiment IV the exercises placed little emphasis on the element of color but much emphasis on the elements, form and notan. The difference in gains between experimental and control groups, however, was greatest for the element, color.

These results, in conjunction with those secured in Experiments I and II, indicate that differences approaching significance in gains between experimental and control groups did occur on sub-tests involving a different element of art than was involved in the experimental exercise series.

From these results, the comparative gain on Experiments III and IV and the comparative gain on sub-tests, the conclusion is drawn that there was actually a change in the ability of the subjects to judge æsthetically between the merits of different pictures and that the increase in æsthetic judgment score was not due to an indirect type of coaching upon the specific type of exercises utilized in the *McAdory* Art Test.

V. Conclusions regarding æsthetic judgment. The experiments thus far presented seem to justify the conclusions that among elementary school children of the second, third, fourth, and fifth grades (1) practice in judging exercises of varying degrees of æsthetic merit without understanding the principles of art involved does not produce a significant increase in æsthetic judgment; (2) a knowledge of the principles of art with practice in the application of these principles to exercises of varying degrees of æsthetic merit will produce a significant increase in æsthetic judgment; (3) although a knowledge of the principles of art with a small amount of practice in the application of these principles will produce an increase in æsthetic judgment, it will not produce as great an increase as a knowledge of the principles of art with a greater amount of practice in the application of these principles.

VI. The effect of the experimental procedures in Experiments III and IV upon the development of æsthetic analysis. The experimental program definitely instructed the subjects as to the manner in which a picture is æsthetically organized. It was desired to determine to what extent the children were able to understand and profit by this training. Accordingly, each child was individually questioned according to a predetermined series of questions which were designed to sample the child's ability to discern the æsthetic organization of Franz Marc's *Red Horses*. The child was thus questioned before and after each experiment. These questions were not a test in the standardized sense. The responses which they provoked should be considered rather as a sample of the child's manner of thinking about a picture, as a painting from a child is considered as a sample of his production. The reservations should also be made that the type of æsthetic analysis which was called forth by these questions is far from being a complete analysis of the æsthetic organization of the picture. Any responses which indicated a consideration of one portion of the picture in relation to the whole were considered as evidence of æsthetic analysis. Thus the questions may more properly be said to have stimulated responses which were indicative of a rudimentary æsthetic analysis. It is assumed, however, that æsthetic analysis is found in many degrees and that such responses indicate a low degree of ability to analyze the æsthetic organization of a picture.

The series of questions is presented below:

This picture is called *Red Horses* and it was painted by an artist named Franz Marc. I am going to ask you some questions to find out if you like the picture and why you like it.

1. Do you like the picture? What do you think is good about it? After the training period, the following question was added: Can you see anything in the picture that we studied about in class? What we studied about in class may help you to answer some of the questions.

2. What is good about the way the artist has painted the picture? You know we say that a person draws something well. What has the artist painted well?

3. Do you like the colors he has used or do you wish that he had used different colors? Why?

4. What do you look at most in the picture? What makes it (or them) show up or stand out so much?

5. Why do you suppose the artist has put blue here and here and here (pointing to water, bush, trees, and valleys) instead of putting blue just down in this corner?

6. Why did the artist make this such a light green—almost yellow—back of the horses instead of making it this dark green up on the hill?

7. How would you have liked it if the artist had made the horses standing on flat ground instead of hills? Would you have liked it as well?

8. Why did the artist put the horses where he did in the picture? Why didn't he want to put the horses farther over on this side in the picture (to the right side)? Why didn't the artist want to put them back here on the hills?

9. I wonder why the artist painted the horses standing that way. You see they are turned different ways and one is a little higher than the others.

10. If the artist had left the middle horse out, would you like the picture as well? Why?

Why do you suppose he made the middle horse a different color?

11. Would you like the picture as well if he had left out this red horse (right) and painted only these two horses in the picture?

12. Why do you suppose he made the bush blue instead of green? How would you have liked it if it had been green?

13. How does the picture make you feel? Why? (If no response was given, the following question was added.) Does it make you feel sad or does it make you feel some other way? How is it different from a sad picture?

14. (After training) Can you find anything else in the picture that we talked about in class?

The responses to these questions were analyzed in the following manner. The entire series of responses was examined and typical answers arranged in a tentative classificatory system. Each paper was read with one type of response, such as balance, in mind. Responses classifiable under this head were recorded in such form that simple counting would give a final score for the groups or sub-groups for which a score was desired, and so that quick reference could be made to the original record from which the response was taken. The first few responses of a sample record are recorded below:

			CONTRAST IN VALUE FOR EMPHASIS
COOPER SCHOOL	C GROUP	TEST III	DARK ON LIGHT—LIGHT ON DARK

1. They are dark and the lighter color makes them show up more (6). Dominance of dark on light, horse's head (14).

2. I suppose the horses show up more. They're dark and it's light (6).

X or C indicates that the group is the experimental or control group. I, II, or III indicates the number of the test in the testing series. The final phrase in the heading characterizes the type of response enumerated below. The number at the beginning of the response is the code number for the individual in the group. The number in parentheses following the response indicates the question to which it is the answer.

The above process was continued until all of the responses had been classified. Those responses which did not fall under any of the classifications originally devised were classified under headings which seemed applicable.

These record sheets were presented to a group of ten judges who were asked to read them determining if the choice of the heading characterizing the response was correct and checking responses which they considered improperly classified. These judges were also asked to write their responses to the same questions asked the children. On the basis of these answers and comments, the classification of the responses was somewhat revised. Each classification was grouped under one of two heads—æsthetic and non-æsthetic. Non-æsthetic responses included unclassifiable responses, responses indicating an incorrect æsthetic response, responses in which the general attitude exhibited seemed to refer to the picture as a substitute-object (naturalistic representation) and æsthetic responses which were so incomplete as to make them meaningless as an indication of the child's knowledge of the principles of art. Æsthetic responses included those in which there seemed to be a partial or complete realization of the picture as an organization of formal relationships. The final classificatory system is presented below:

Categories for the Classification of Æsthetic Analysis Responses

A. *Non-æsthetic.* Responses which are unclassifiable, which are incorrect æsthetically, or which refer to the subject matter of the picture rather than to its æsthetic organization.

Unclassifiable responses. Ambiguous character, not assignable to any category. Examples: "Some pictures show all hills, I have an idea."

Incorrect responses. Affirmative answer to a question which involves an undesirable change in the composition. Example: " It would look just as well with blue just in this corner."

Interest in subject-matter. Interest in the object portrayed as an object, in the activities of the object, and in the setting of these activities in time and place. Object not considered as a part of the pictorial organization. Examples: " Yes. I like stories about horses in pictures." " The horses look like they are going to be captured." " Water. The horses could drink it." " Looks about like sunset."

Enumeration. Naming over or description of objects in the picture without further elaboration. Examples: " Rocks stick up more along here." " Hills. Horses. Bushes."

Naturalism. Desire for photographic realism. Examples: " Horses aren't very pretty for their redness. Sort of tan better for horses." " Wanted them (horses) that way. Saw them that way." " Horses don't stand straight all the time. These look natural."

Logical responses. Adequate from a reasoning but not from an æsthetic standpoint. Examples: " Shows it's warm because the grass is green." " Bush is blue. Flowers maybe." " That's where the horses ate off of it (grass)."

Drawing. Applications of what has been learned in the drawing class dealing with proportion, naturalistic representation, and skill in handling the media, but failure to relate the elements to the composition as a whole. Examples: " Way he drew the horses. You can't always get the right proportions." " Colors don't run together. It's not messy." " Didn't take too many pains painting it. He just sketched it." " It would be harder to draw if the horses were up on the hill."

Preference for object variety. Preference for variety of ways of portraying an object without reference to æsthetic relationships. Examples: " So you could see the back of the horses." " Wanted to get all three of them. This horse may be half-way in, half-way out." " Horses' shapes. These two looking at you. Others not. Gives a better view of his neck."

Generalities. Indicative of an evaluation but not indicative of ability to analyze the basis for that evaluation. Examples. " That's pretty." " Horses arranged nice." " Probably liked it better." " Picked good colors for the sky and ground."

Stereotyped phrases—color. Phrases of common occurrence indicative of a preference for certain color combinations which do not evidence a real comprehension of the problems involved in such combinations. Examples: " The colors are blended in." " Look like they match." " Color goes good together."

Preference for the unusual. Preference for a subject or technique because it differs from that usually found. Examples: " Bushes. Look so unordinary." " Purple in the corner. Don't see so much of that in pictures." " So many pictures have level ground."

Personal associations. Projection of the observer within the frame of the picture as participating in the activity portrayed. Examples: " Be on a horse's back riding. Never been around a whole lot of hills. Would be on a lot of hills." " Like to get outside and play."

Sensuous enjoyment of color. Enjoyment of a particular color for its own sake without reference to the remainder of the composition. Examples: " Like the background. Colors he colored." " Yes. The red horses. My favorite color." " Pretty. Like light colors better than dark."

Perception of movement. Perception of the portrayed movement of an object but not movement inherent in the pictorial relationship (*e.g.,* rhythm).

Examples: "Like the action of the horses. The way they are standing." "Like the way it's painted. Like how the horses are kind of stamping around."

Perception of texture. Identification of texture but failure to relate it to the remainder of the composition. Examples: "Tails. Bushy." "Horses smooth." "Texture. Stream rough."

B. *Æsthetic.* Responses indicative of (1) the arrangement of objects within the picture space to facilitate a unified impression; (2) the relative emphasis placed on different objects to facilitate grouping and sub-grouping within the picture; (3) the organization of the picture in the third dimension; (4) the selection of colors, forms, and values related to the whole (harmony); (5) the employment of sufficient contrast within the unitary organization to increase the interest of the picture (variety); (6) the use of pictorial elements to express mood.

Space filling. Realization that each portion of the picture space must enhance the total composition. Examples: "To fill the picture out more." "It makes a difference in the landscape. If nothing in between have to have another hill or something." "It would leave that space empty. Wouldn't look good."

Distribution in space. Realization of the necessity of proper distribution of elements within the picture space. Examples: "Too crowded (if on one side)." "Don't have all just up together in one place." "Don't look good to have too much in a picture." "Too small for the horses. Have to make it a bigger picture."

Grouping. Unification by placement of shapes within groups. Examples: "Prettier in a group than if all over the picture." "Positions. If apart be three captains. If together just one."

Pattern of color. Balanced distribution of color over the entire picture space. Examples: "Look funny, blue all in one place." "Looks nicer to have them scattered around like a patchwork quilt."

Repetition. Relating different portions of the picture by use of color, shape, or value in each portion. Examples: "Color alike. Blue up here and down here." "No. It seems if two of one thing, want to have it the same kind." "Color would not look like the other side."

Sequence. Unification of a number of objects by progression in size. Examples: "Sequence. Low head, higher, top one way up high." "Like these. They make a sequence of color. Light, dark, darker." "Sequence. The hills. They get bigger as they go back."

Radiation. Unification of a number of objects into a larger shape characterized by radiating form. Examples: "The horses make a kind of radiation. They go out different ways." "They look better. One head is going in each direction."

Geometrical form. Unification of a number of objects into a larger form such as a square, triangle, circle, S-curve, etc. Examples: "Makes a triangle—horses." "Mountain a circle." "Horses make a circle." "Bush. Like it's round. Hills are round." "S-curve. Head of horse down through neck of other."

Transitional line. Relation of the oblong picture shape to the general visual area. Example: "The hills round off the corners." "Transitional line. Ravine." "So look like transitional around. Circular top."

Direction. Relation of objects by colors, values, or shapes leading from one to the other. Examples: "These horses sort of point to the middle one. Captain. Dominance." "One pointing. The horse down to the ravine." "Yellow back there. See it first, then look at the horses." "Blue in the corner. Look at it, then notice them."

Balance. Equilibration of elements in the picture space. Examples: "Would have been too much on this side. The picture wouldn't balance right." "The horses wanted balance—wanted hills." "Top-heavy."

Emphasis. Variation in stress to provide variation of interest-value. Examples: "No. There would be just the hill there. There would be too much hill." "It makes the horse stand out too much."

Domination. Recognition of one portion of the picture as being most important but no recognition of the means employed to make it so. Examples: "Horses. Supposed to be the main thing." "Horses and boulders center-piece. Boulders and others background."

> *Domination by isolation.* Examples: "Only thing in there except ground. Attract attention more." "Horses the main things. No more other things you want to look at."
>
> *Domination by size.* Examples: "Horses. Main object. Bigger." "No. Horse takes up more room. Not well if don't have enough of what painting about." "Got captain. Dominance. Dark red horse. Up higher."
>
> *Domination by bright color.* Examples: "Color of horses makes them stand out." "Brightest color." "Background not so bright as other."
>
> *Domination by different color.* Examples: "Mostly green. Have to make it blue to show up." "If background the same color it wouldn't show up. Different color."
>
> *Domination by different value.* Examples: "Colors. The rest of the colors lighter. Horses darker." "Dark green wouldn't look right against dark red horses." "Around the horses, colors are lighter than others."
>
> *Domination by position in the center.* Examples: "That was his main topic, in the middle." "Horses. Right in the center. See them better because first things seen."
>
> *Domination by position in the foreground.* Examples: "Horses are more important in the picture so he put them up in front." "Because they would show up better in front than in back."
>
> *Domination through movement.* Examples: "It is the most moving creature." "Attract attention. Prancing around." "Action makes them show up."
>
> *Domination by framing.* Examples: "They have a sort of border on both sides." "Framing around the horses." "Helps surround the picture."

Depth. Realization of the presence of three-dimensional relationships. Example: "Hills. Really look like back of each other."

Rhythm. Movement through the relationships of pictorial elements. Examples: "Look kind of rolly and fluffy. Like the waves. Flow like that." "Rhythm. Goes up horses' backs."

Color harmony. Colors considered in terms of mutual relationship and not in terms of preference for isolated colors.

> *Color harmony through the use of values of the same color or through the use of few colors.* Examples: "If too many colors not look good." "All the horses are different shades of red."

> *Color harmony through the use of colors which are not contrasting colors.* Examples: "Green next to red is not as good as blue." "Wouldn't want blue and yellow, rather than blue and red."
>
> *Color harmony through the use of colors containing a common hue.* Examples: "Colors go good together. Pink and reddish pink." "Thought red and purplish and blue went good together." "Colors matching. All have been darkened by a dark color." "Yellow will divide into all the horses."
>
> *Color harmony. Responses indicating that it is not obtained by the use of contrasting color.* Examples: "If green it wouldn't go good with the horses." "Red and green opposite. Together they would not look very well."

Variety of color. Preference for different colors, a dislike of too much of one color, a desire for more of a color. Examples: "Better kinda yellow or brown but then, if all the same color, it would be rather dull." "The middle horse is not the same color. You wouldn't want the picture all the same color." "It makes it prettier. So many colors in it."

Variety of color through contrast in color. Examples: "Lot of green and a little red. Makes the horses show up."

Harmony of value. Examples: "All sort of light so they go in good together." "Next to the horses. Because the horses are bright." "If there was a real dark background and the horses were real light it would look funny."

Variety of value. Examples: "Not put the sun in too many light colors. A little light and a little dark are better than all light." "The hills bring my attention. Some places are dark and others real light."

Harmony of Size. Examples: "They (horses) are all the same size." "The hills go better with the horses. The horses are big and high. The hills are about level with the horses—to the top of the horses."

Variety of size. Examples: "You wouldn't want them all the same size." "The hills. They are not all the same size."

Harmony of shape. Examples: "Harmony of shape. The horses are all the same shape. The ground. The hills." "If hills were squares and the horses curved like—squares and curves don't go together." "Everything is like a circle. Round. Rock. Hill. Horses." "Yes. How the land is painted. Not like real land. More like the way the horses are standing."

Variety of shape. Examples: "You wouldn't want them all standing the same way. One has his head down, one has his head turned." "Just makes it look better." "Not the same thing over and over like a cornfield."

Mood through color. Use of color as a medium to express mood. Examples: "Happy feeling. The way it's colored. The colors they use in it. All the colors are joyous colors—real bright." "Could have mixed a little brown with the horses, yet not so good, so joyful, that way." "Gay and happy, full of pep. Bright color."

Mood through movement. Use of movement as a medium to express quality of mood. Examples: "Probably wanted to show them peppy and raring to go. The way they are looking and the way their feet are." "Jumpy. Horses so lively and everything. Want to turn somersaults. Looking different ways. Running around."

Mood through rhythm. Use of rhythm as a medium to express mood. Examples: "Happy. Lines curved. When sad, straight." "Foxy. Hills look so good and bumpy." "Hills look like flying—joyful." "Looked stale. Has rhythm up there in the hills. Makes it feel gay and happy."

More inclusive harmonies. Relating of one type of element to another type. Example: "The colors fit what they (horses) are doing."

The child's æsthetic analysis score was derived by determining the proportion of the total number of æsthetic responses to the total number of responses made by the child. The significance of the difference between initial and final mean scores for each group and between the final scores of the two groups were obtained in the manner described on page 6. The results are presented in Table VIII.

In Experiment III each group gained a significant amount between initial and final tests, and the difference between the final scores of the experimental and the control groups was significant. In Experiment IV each group gained a significant amount between initial and final tests but the difference between the final mean scores of the two groups was not significant. The initial score of the experimental group was much lower, however, than that of the control group. Accordingly the significance of the difference between the gains of the two groups was computed and is presented in Table IX.

The chances that the gain in Experiment IV was significant proved to be as great as those in Experiment III. These results indicate that training in the principles of art does definitely increase the child's ability, not only to understand an analysis of a picture, but to apply correctly the same type of analysis to a new picture situation.

VII. The effect of the experimental procedure upon the approach to the picture. The scores on the compositional analysis responses may be used to check the validity of the hypothesis that, accompanying an increase in æsthetic judgment and analysis scores, there is a decrease in the number of responses indicating that the picture is considered as a substitute-object, and an increase in the number of responses indicating that the picture is considered as an æsthetic organization. The significant increase in mean scores on the *McAdory* Art Test and on the æsthetic analysis responses indicated an increase in these two abilities. The changes in mean score in certain categories of the æsthetic analysis classification were investigated to determine if

TABLE VIII. *Significance of the difference between initial and final scores of experimental and control groups on ability to analyze verbally the æsthetic merit of a painting and between final scores of experimental and control groups for Experiments III and IV*

			First Test			Second Test			Between First and Second Test			Between Second Tests of each Group		
Experiment	Group	No. of subjects	Mean	S.D.	P.E._{av}	Mean	S.D.	P.E._{av}	r	$\text{P.E.}_{\text{diff.}}$	Critical ratio	$\text{P.E.}_{\text{diff.}}$	Critical ratio	P
III	E	40	49.00	14.55	1.55	70.00	14.14	1.51	.26	2.16	16.67			
	C	41	46.95	16.45	1.73	60.85	16.30	1.72	.44	2.44	7.60	2.29	4.00	99.7
IV	E	39	61.15	15.40	1.66	75.63	11.46	1.24	.44	2.07	9.16			
	C	38	71.47	13.93	1.53	78.54	8.94	.98	.39	1.82	4.84	1.58	1.84	89.0

TABLE IX. *The significance of the difference between the gains in æsthetic analysis scores of the experimental and control groups in Experiments III and IV*

	Group X				Group C						
Experiment	No.	Mean	S.D.	P.E._{av}	No.	Mean	S.D.	P.E._{av}	$\text{P.E.}_{\text{diff.}}$	Critical ratio	P
III	40	20.38	18.80	2.00	41	13.30	14.15	1.49	2.50	2.83	97.0
IV	39	13.95	15.15	1.63	38	6.30	13.05	1.43	2.17	3.53	99.0

a change had actually occurred in the manner in which the child set about his examination of the picture. The mean number of responses for each category for series one, two, and three, of the administration of the questions to Groups A and B is presented in Table X.

The significant categories to examine in this respect are those of interest in subject-matter, enumeration, naturalism, logical response, and drawing. In Experiment III, following the experimental period, there is a decrease in each of these categories with the exception of drawing. The mean change in score when the scores for these five categories were combined was —.34. In Experiment IV, following the experimental period, the mean number of responses decreased in the categories, interest in subject-matter, enumeration, and drawing, and increased in the categories, naturalism and logical response. The change in mean score of these five categories combined, however, was —.16. Even greater than the decrease in mean scores was the decrease in the proportion of such responses to the total number of responses, since the decrease on non-æsthetic responses was accompanied by a large increase in the number of æsthetic responses.

Probably, due to the nature of the test situation, the æsthetic analysis responses continued to be based to a greater extent on the type of responses characterized as substitute-object responses than in the case of the æsthetic judgment responses. In the æsthetic analysis test-situation, a type of subject-matter, of very great interest to children of this age and accordingly likely to arouse many associations from past experience, was portrayed in such a manner that photographic representation was flagrantly violated. In contradistinction, the æsthetic judgment test consisted of a large variety of types of subject-matter, many of which are not inherently interesting to children of this age and was presented in such form that often alternative choices were equally acceptable photographic representations. Accordingly, although a direct inference from one situation to the other may not be valid, it does appear probable that in the æsthetic judgment test as in the æsthetic analysis responses fewer judgments

TABLE X. *Mean scores for Groups A and B on each type of æsthetic analysis response*

	Group A			Group B		
Category	Test 1	Test 2	Test 3	Test 1	Test 2	Test 3
Non-æsthetic:						
Unclassifiable	.02	.03	.00	.10	.02	.00
Incorrect	.04	.00	.05	.14	.12	.02
Subject-matter	2.42	1.78	1.56	2.34	2.41	1.72
Enumeration	.31	.23	.28	.24	.17	.13
Naturalism	1.26	.75	.95	1.04	.85	1.04
Logical responses	1.44	1.05	1.26	1.18	1.41	1.59
Drawing	.33	.58	.35	.40	.41	.17
Object variety	.16	.25	.35	.10	.22	.24
Generalities	.78	.56	.23	.78	.56	.41
Stereotypes, color	.49	.65	.51	.78	.56	.50
Preference for the unusual	.07	.05	.09	.02	.10	.11
Personal associations	.09	.15	.05	.08	.12	.04
Enjoyment color	1.22	1.20	.91	1.36	1.41	.65
Perception movement	.29	.13	.16	.32	.24	.22
Perception texture	.07	.10	.19	.02	.15	.06
Æsthetic:						
Space filling	.42	.98	.84	.70	1.22	1.17
Distribution in space	.02	.05	.21	.00	.07	.26
Grouping	.16	.08	.05	.04	.24	.11
Pattern of color	.29	.58	.47	.32	.32	.33
Repetition	.44	.68	.53	.26	.37	1.15
Sequence	.02	.20	.35	.02	.02	.93
Radiation	.00	.00	.23	.04	.02	.06
Geometrical form	.02	.23	1.33	.02	.05	1.28
Transitional line	.00	.13	.16	.00	.00	.02
Directional line	.13	.40	.47	.00	.12	.50
Balance	.44	1.78	3.70	.80	1.51	3.28
Emphasis	.11	.10	.14	.08	.17	.13
Domination, general	.11	.28	.23	.04	.12	.24
Isolation	.07	.05	.02	.00	.02	.00
Size	.67	1.18	1.67	.82	.80	1.13
Bright color	.62	.90	.95	.60	.66	.85
Different color	.87	1.68	2.00	1.06	1.59	2.13
Contrasting color	.02	.15	.14	.00	.07	.11
Different value	.18	.95	1.51	.26	.76	1.96
Center position	.36	.98	1.02	.52	.66	.87
Foreground	.64	.93	1.12	.58	.73	.87
Movement	.11	.18	.21	.10	.15	.37
Framing	.00	.03	.14	.02	.00	.06
Depth	.22	.28	.33	.46	.37	.37
Rhythm	.22	.68	.91	.30	.29	.37
Color harmony						
Through use of values of the same color	.00	.05	.19	.02	.05	.37
Through use of colors not contrasting	.00	.10	.07	.02	.02	.13
Through use of colors containing a common color	.00	.15	.37	.00	.12	.04
Not contrast	.02	.15	.12	.04	.07	.41
Color variation	1.16	1.48	2.26	.52	1.56	.06
Color variation through contrast	.00	.05	.12	.40	.07	1.89

TABLE X—*Continued*

Harmony of value	.04	.28	.21	.02	.00	.00
Variety of value	.09	.10	.16	.08	.07	.06
Harmony of size and shape	.02	.00	.05	.00	.02	.11
Variety of size and shape	.13	.48	.40	.16	.39	.30
Mood through color	.44	1.13	1.30	.38	.49	.63
Mood through movement	.96	1.15	1.23	.62	1.17	1.20
Mood through rhythm	.02	.13	.09	.06	.00	.09
Harmony of major qualities	.00	.08	.07	.00	.00	.04

were based upon such criteria as subject-interest, photographic representation, and technical ability, and more judgments were based upon the criteria of the principles of art following the training period than before, and that the increase in æsthetic judgment score was accompanied by this change in criteria. As the emphasis in the experiment was placed upon the explanation of positive methods of organizing the picture and not upon the inadequacy of substitute-object criteria, the results seem to confirm to a certain extent the assumption made at the beginning of this report that the two are mutually exclusive.

VIII. The effect of the experimental conditions upon the development of the ability to produce an æsthetically meritorious composition. The ability of the child to organize an original painting according to æsthetic relationships was also tested before and after Experiment III and after Experiment IV. The children were tested in groups of six or eight children at a time. Each child was provided with an easel, brush, and paper. A large supply of various paints was placed on a table for the use of all of the children. After the children had been assembled the following directions were given:

> We are going to paint a picture today. First, I will tell you how to use these paints. You dip your brush in a glass of water, and wipe it off carefully on the edge of the glass so that there is not too much water in the brush. If there is too much water the paint will run down your picture. Then dip just the tip of the brush into the paint jar and start to paint. If the color is too dark, put more water on your brush. If the color is too light, do not put so much water on your brush. You may take one jar of paint at a time. As soon as you have finished with that color, bring it back to the table and take the next color of paint that you wish to use.
>
> Now let me tell you what I want you to paint. You may paint an illustration of some story you have been reading such as William Tell or Robinson Crusoe, or you may paint a picture about winter sports or some other subject that you like. I want you to put at least two objects in your picture, and to put in a

background; that is, to put in hills, or trees, or water, or whatever you would like for a background for your main objects.

As we have only a short time to paint we will have to make a sketch. Have you made sketches in school? When you sketch anything, you paint in the main objects of your picture quickly and do not try to paint every little thing in the picture. Put in the main objects first, and then fill in the background.

I will give you a minute or two to decide what you would like to paint. Then decide which color you will need and take a jar of that color of paint to your easel and begin. Remember to put in at least two objects first and then fill in the background. You will find all of the other things you need for painting at your easel. You will have fifteen minutes in which to paint your picture.

All of the paintings, including three for each child, were judged at one time. The judges were ten research assistants in the Psychology of Art division at the University of Iowa. These judges were artists, or were individuals possessing considerable experience in judging children's paintings. They represented a wide variety of points of view and training. The following instructions were given for the ranking of the paintings.

These paintings were made under the following conditions. The children were allowed 15 min. in which to paint. They could paint any subject-matter they desired although illustrations of stories such as William Tell or Robinson Crusoe, or the subject "Winter Sports" were suggested. Two conditions must be fulfilled: (1) there must be at least two objects in the picture, and (2) there must be a background. They were prohibited from painting anything in the room.

Due to the limitation in time and to the fact that water color was a new medium to the children, the paintings are necessarily very rough. Accordingly, it may be necessary to be quite generous in your judgments in order to differentiate the paintings sufficiently. Disregard skill in technique and representation and base your judgments upon the relative attainment of composition, involving rhythm, balance, unity, harmony, *etc.* Any consideration of what a fifth-grade child *should* do should likewise be disregarded and judgments based entirely upon the relative merit of the compositions within this group.

You are to group the paintings into nine piles, representing nine degrees of merit from very poor through a mediocre performance to very good. Reserve the lowest pile for pictures entirely lacking in composition, the highest pile for pictures you consider to exhibit outstanding merit. It may be helpful to separate the pictures into three piles and later to sub-divide each of these piles into three degrees of merit. After you have completed your ranking, go through the pictures again and again, redistributing the pictures until you are relatively sure that further examination would not change the placements to any great extent.

When you have completed the ranking process, tabulate the paintings within each pile by recording the number stamped on the back of the picture. The group representing the lowest degree of merit will be designated Group 1, that representing the highest degree of merit, Group 9.

When the recording has been completed, shuffle all the pictures together so that no trace of your ranking remains.

The final score for each painting was in the form of an average of the ranking of all the judges. As there was no suitable scale by which the paintings could be rated there were distinct differences in the form of the distributions.[13] In combining the judgments, the scores for each judge were transmuted into sigma values on a normal distribution and these values were averaged to give the final score of a painting (**15**, pp. 155–160). The reliability of these judgments, as measured by correlating the average of five judges with the average of the remaining five judges, was .62±.03. The reliability of the final score, as determined by the average of the scores of the ten judges, was determined by the application of the Spearman-Brown prophecy formula for double the number of judgments to the above correlation. The resulting value was .77.

The significance of the difference between initial and final scores of each group, and of the difference between the final scores of the two groups was determined as before. The results are presented in Table XI. These results fail to indicate that a significant difference occurred either between initial and final tests within each group or between the final tests of the X and C groups. When the significance of the difference between the gains, shown in Table XII, was computed the probabilities seem sufficiently high to indicate that the experimental procedure did produce a change in the ability of the child successfully to organize a composition. Apparently the experimental conditions were detrimental in their influence. However, it is probable that practice is fully as important in production ability as in æsthetic judgment ability. The experiments provided for a large amount of practice in the recognition of the principles of art. It is recognition that is involved in æsthetic judgment and analysis. There was, however, no opportunity provided in which the child might attempt to apply the principles in the composition of his own paintings. Without such opportunity for practice in application, it is possible that an abrupt attempt at such application within the very short space of time allotted for the production of the painting confused the child to such an extent that a decrease rather than an increase in the quality of the composition resulted.

---

[13] Due to the sketchiness of the paintings the Iowa Drawing Scale could not be used, as the paintings all tended to fall into the lower classes so that sufficient discrimination could not be obtained.

IX. The permanence of the increase in æsthetic judgment score produced by Experiment II. As both groups of children received training in Experiments III and IV a retest some time after the conclusion of these experiments was not expedient in that there was no standard by which to compare their final performance. In Experiment I, however, the training of the A group had resembled so closely in nature and in results a simple repetition of the test, that a retest of the groups used in Experiment I and II was feasible. Accordingly a retest of both groups was made approximately six months following the conclusion of Experiment II to determine the stability of the increase in æsthetic judgment produced by the experimental procedures.

The results, presented in Table XIII, indicate that the difference between the two groups had now decreased to 1.11 and was no longer significant. It appeared that the understanding as to the proper criteria for æsthetic judgment and analysis produced by the experimental techniques was not sufficient to continue in operation during the intervening summer vacation period and the activities of the subsequent school year, neither of which involved this type of training. On the other hand, the gain of the control group is probably that of the natural increment, that is, the increase in score from grade to grade characteristic of the McAdory Art Test. Although there are no data upon which to base any definite conclusions as to what actually occurred to bring about these results, the theoretical explanation is advanced that unless a situation is provided in which a sufficient opportunity is furnished for continuous practice in the recognition of these principles, failure to progress to a significant degree in the ability to judge æsthetically ensues.

X. Summary and conclusions. This investigation was designed to determine the effect of two factors, direction as to the proper criteria of art and practice in the perception of manifestations of these criteria, *e.g.,* in the development of æsthetic judgment,

TABLE XI. *The significance of the difference between initial and final scores of experimental and control groups and between final scores of experimental and control groups for Experiments III and IV*

Experiment	Group	No. of subjects	First Test			Second Test			Between First and Second Test			Between Second Tests of Each Group		
			Mean	S.D.	P.E.$_{av.}$	Mean	S.D.	P.E.$_{av.}$	r	P.E.$_{diff.}$	Critical ratio	P.E.$_{diff.}$	Critical ratio	P
III	E	40	.11	.52	.055	.02	.53	.056	.64	.048	1.88	.079	.63	66.0
	C	38	—.10	.50	.055	.07	.47	.051	.32	.062	2.74			
IV	E	33	.02	.42	.049	—.02	.46	.065	.55	.080	.80	.078	3.85	99.5
	C	38	.08	.58	.063	.28	.43	.048	.04	.080	2.50			

TABLE XII. *The significance of the difference between the gains in the ability to produce an æsthetically meritorious composition of the experimental and control groups in Experiments III and IV*

Experiment	Group X				Group C				P.E.$_{diff.}$	Critical ratio	P
	No. of subjects	Mean gain	S.D.	P.E.$_{av.}$	Number	Mean gain	S.D.	P.E.$_{av.}$			
III	40	—.08	.45	.05	37	.21	.59	.07	.086	3.37	98.8
IV	33	—.04	.59	.07	37	.21	.78	.09	.111	2.23	93.0

TABLE XIII. *The significance of the difference between the final McAdory Art Test scores of experimental and control groups six months following the expiration of Experiment II*

Experiment	Group	No. of subjects	First Test			Second Test			Between First and Second Test			Between Second Tests of Each Group		
			Mean	S.D.	P.E.$_{av}$	Mean	S.D.	P.E.$_{av.}$	r	P.E.$_{diff.}$	Critical ratio	P.E.$_{diff.}$	Critical ratio	P
II	E	34	43.32	8.34	.965	44.22	8.64	1.000	.87	.50	1.80			
	C	41	38.94	9.48	1.000	43.11	8.64	.911	.83	.57	7.32	1.35	.82	71

æsthetic analysis, and the production of an æsthetically meritorious composition. The subjects were children of the second, third, fourth, and fifth grades in two typical Iowa schools. These subjects were divided into two matched groups on the basis of grade, sex, IQ, MA, CA, and *McAdory* Art Test scores. The effect of a knowledge of the principles of art and of practice in the recognition of applications of these principles was tested by a series of four experiments. In Experiments I and II the subjects were tested before and after each experiment by a modified form of the *McAdory* Art Test and by two subsidiary tests: an arrangement of items from the *Whorley* Unity Scale, and the *Lewerenz* Test in Fundamental Abilities of Visual Art, Part I, Test 1. In Experiments III and IV the subjects were tested before and after each experiment by the standardized form of the *McAdory* Art Test, by a series of questions requiring an æsthetic analysis of Franz Marc's painting *Red Horses,* and by the production of an original composition.

In Experiments I and II the experimental groups were presented with a series of paired or multiple-choice exercises, one member of which exemplified, the other of which violated, a principle of art. In Experiment I the subjects were required to judge upon the relative merit of the members of each exercise but were not provided with any explanation of the factors upon which the judgments should be based. In Experiment II the subjects were required to judge as to the relative merit of each portion of the exercise, as in Experiment I, but they were also directed as to the factor upon which the judgment should be based by instruction as to the principle of art which was involved. In Experiment I no significant difference between the gains of the experimental and control groups occurred, while in Experiment II a significant difference did occur. Variations in Experiment II indicated that the gain in æsthetic judgment score was also dependent upon the factor of practice in the recognition of the principles, since a group having judged more than 90 exercises gained in æsthetic judgment score to a greater extent than a group having judged fewer than 90 exercises.

To determine whether increases in æsthetic judgment score in Experiment II were a function of knowledge of and practice in the recognition of the principles of art, and not a function of coaching upon exercises similar to the test exercises, a comparison was made of the gains upon sub-tests involving different art elements. A significant increase in æsthetic judgment score occurred upon portions of the test situation which dealt with applications of the principles to elements not similar to those involved in the experimental exercises.

In Experiments III and IV the groups were presented with the same explanations as to the principles of art involved in the exercises. In Experiment III the subjects were required to abstract a common principle from a number of illustrations which varied in many other respects. In Experiment IV the subjects were required to note the presence of a certain principle of art in one member of the exercise and the absence of this principle in the other member in order to judge as to their respective merits. In Experiment III, although the exercises differed considerably from the exercises involved in the test situation, the significance of the difference in gain between experimental and control groups was approximately the same as that in Experiment IV in which the exercises were more similar to the test exercises. Cultivation of a comprehension and recognition of the principles of art therefore seems to bring about an increase in æsthetic judgment regardless of the type of exercises which may be utilized to illustrate the principles.

An analysis of the responses to the picture *Red Horses* demonstrated that, following the experimental period, the subjects possessed a much greater understanding of what is involved in the analysis and judgment of æsthetic merit than before and that, accompanying an increase in the number of responses indicating a comprehension of the picture as an organization of æsthetic relationships, the number of responses indicating a comprehension of the picture as a substitute-object decreased in number and in proportion to the total number of responses.

Significance of the differences in mean ratings of the paintings produced before and after the experimental procedure did not indicate an increase in production ability similar to that secured in æsthetic judgment and analysis abilities. There was apparently a tendency for the experimental group to lose slightly in mean score and for the control group to gain. Failure of the experimental procedure to provide opportunities for application of the principles to production may account for the result. For an adequate understanding of the effect of a knowledge of the principles of art upon production, a more refined experimental procedure, accompanied by opportunities for application, is needed.

The following conclusions seem to be justified by the foregoing results: (1) The principles of art can be explained in such a manner that they can be understood and applied by children in the second, third, fourth, and fifth grades. (2) An understanding of the principles of art produces a significant increase in the ability to judge between varying degrees of merit. (3) An understanding of the principles of art produces a significant increase in the ability to analyze the æsthetic organization of a picture. (4) Accompanying a significant increase in the æsthetic judgment and the æsthetic analysis scores a decrease occurs in the comprehension of the picture as a substitute-object and an increase occurs in the comprehension of the picture as an æsthetic organization. (5) An explanation of a principle of art accompanied by illustrations exemplifying its application in many ways brings about a greater increase in æsthetic judgment than an explanation accompanied by illustrations exemplifying its application in but few ways. (6) An understanding of the principles of art without opportunity for the application of these principles in the production of art is not sufficient to increase the ability of the child to organize an æsthetically meritorious composition, as such ability was measured in this experiment. (7) With complete cessation of the type of activities provided by the experimental procedure over a considerable period of time, there is failure to progress to a sufficient extent so that the experimental group retains a significant superiority over the control group.

Bibliography

1. Christensen, E. O., and Karwoski, T. F. A test for art appreciation. *J. Educ. Psychol.*, 17, 1926, 187–194.
2. Garrett, H. E. Statistics in psychology and education. N. Y., Longmans Green, 1926.
3. Goldstein, H. and V. Art in every day life (2nd ed.). N. Y., Macmillan, 1932.
4. Jacobson, W. E. An experimental investigation of the basic aesthetic factors in costume design. *Psychol. Monog.*, 45, 1933, 147–184.
5. Kuklmann, F., and Anderson, R. G. Intelligence tests. Instruction manual (4th ed.). Minneapolis, Educ. Test Bur., 1933.
6. Lewerenz, A. S. Scientific measurement in the realm of art. [In] Yrbk. of the Southwestern Educ. Res. and Guidance Assn. Los Angeles, Research Service, 1927.
7. Lewerenz, A. S. Tests in fundamental abilities of visual art. Los Angeles, Research Service, 1927.
8. Lindquist, E. F. The significance of a difference between "matched" groups. *J. Educ. Psychol.*, 22, 1931, 197–204.
9. Maier, Norman R. F., and Reninger, H. W. A psychological approach to literary criticism. N. Y., Appleton, 1933.
10. McAdory, M. Art Test. N. Y., Bur. of Publ., Teach. Coll., Columbia Univ., 1929.
11. McAdory, M. The construction and validation of an art test. N. Y., Bur. of Publ., Teach. Coll., Columbia Univ., 1929.
12. Meier, N. C. The Meier-Seashore art judgment test. Iowa City, Bur. of Educ. Research and Service, Univ. of Iowa, 1929.
13. Meier, N. C. The Meier-Seashore art judgment test. Examiner's manual. Bur. of Educ. Research and Service, Univ. of Iowa, 1929.
14. Siceloff, M. M., and Woodyard, E., *et al.* Validity and standardization of the McAdory art test. N. Y., Bur. of Publ., Teach. Coll., Columbia Univ., 1933.
15. Thurstone, L. L. The fundamentals of statistics. N. Y., Macmillan, 1927.
16. Whitford, W. G. An introduction to art education. N. Y., Appleton, 1920.
17. Whorley, K. An experimental investigation of the sensitivity of children to compositional unity. *Psychol. Monog.*, 45, 1933, No. 1, 26–45.

EMPATHIC RESPONSES IN CHILDREN [1]

by

WILLIAM E. WALTON

I. Introduction. Previous studies (6, 9, 12, 14) of empathy, as it is concerned with graphic art, have been made on adult subjects. This investigation was planned to explore the child range from kindergarten through the twelfth grade to discover its genetic development and individual differences at these age levels.[2] In order to interpret children's responses adequately, responses were also secured from adults. It was believed that such a study might furnish additional information concerning (1) the nature of the empathic response, (2) the differences in response between artistic and non-artistic children, and (3) the relationship between the usual lines and colors and facial forms of expression.[3]

Empathy is expressed usually as a process of "feeling oneself into" the perceptual object. Physiological responses, probably visceral and kinæsthetic in nature, take place. However, a lack of recognition of their physiological character causes identification of these empathic responses with the object. As *Langfeld* cites, "The perception obtained by the roundness of an object, or other factors of shape, depends upon similar processes. When we notice the smooth curves of a marble torso, we can probably, if we observe carefully, get a fleeting image of our hands moving in imagination around the figure" (12). We speak of lines as being 'sad,' 'merry,' 'angry.' Empathy is not to be confused with associations which are purely mental. A child, for example, may not like red because the maid in the family wears the color (16). Empathy is closely related to mimetic responses which are responses made in imitation, either directly or through analogy, of some associated quality. 'Loud' colors, for example, are usually associated with noisy individuals, colors decorating noisy places and, in general, unrefinement and crudeness (1, 3, 4). It is believed that empathy is related to artistic ability in that the more artistically talented individual is the one who can more readily and completely identify himself with the object or situation. *Langfeld* (13), *Chandler* (14), and *Birkhof* (2) regard empathy as a factor in æsthetic responses, such as rhythm and balance. Because of the importance of empathic use of lines in cartooning and in the judgment of facial expressions, the writer also included a study of the judgment of facial expression.

---

[1] An investigation in the research program, Genetic Studies in Artistic Capacity, directed by Dr. Norman C. Meier and sponsored by the Carnegie Foundation for the Advancement of Teaching with financial aid from the Carnegie Corporation of New York. Later sections of this study were completed at the University of Nebraska.

[2] This investigation is related to earlier studies reported in *Psychol. Monog.*, 45, 1933, No. 1.

[3] The Ruckmick series of facial expressions were used in part for this portion of the investigation.

II. Technique: construction of tests; preliminary study. On the assumption that test items used by previous experimenters upon adult subjects were too difficult for kindergarten children—as having the child draw a 'merry line'—a simple dual choice method was adopted instead in which the child would designate which line he thinks is 'merry.' Ninety-three dual choice situations with accompanying questions were formulated. To insure comprehension at the first grade level, all words used in the questions were checked against the *Horn* word-list.[4] The situations included lines, forms, and colors and generally recognized principles of art presented in simple line drawings. Four situations are presented in Fig. 1. The child's score was the number of responses made correctly. The correct choice was determined by the judgments of a group of colleagues in the field of the psychology of art.

The test proved too easy, since the preliminary group was approximately 93 per cent correct in its responses, necessitating the use of more difficult materials and a technique which would call for a more precise use of the empathic factor.

---

[4] Horn, M. D. A study of the vocabulary of children before entering the first grade. Baltimore, International Kindergarten Union, 1928.

The final form of the test included four parts. The first called for the child's response to adjectives by drawing lines descriptive of the adjective. In the second, test lines and colors were shown and the subjects were required to select from pairs of words the word having the same meaning as the presented line or color. Part three called for spontaneous verbalization in response to lines and colors. Test four employed the *Ruckmick* pictures and followed the technique of *Gates* (1).

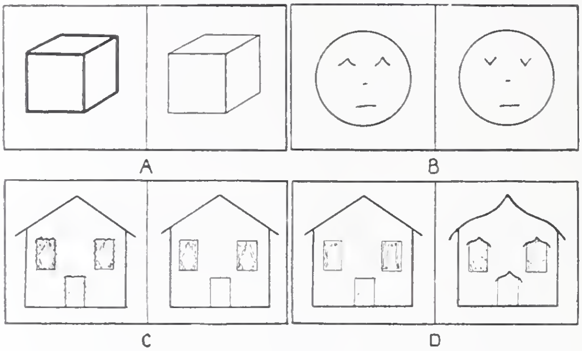


FIG. 1. Examples of items used in preliminary tests.

Questions used with the items shown above:

A. Which box do you think is the heavier?
B. Which boy do you think is feeling good?
C. In which house do you think the cross man lives?
D. In which house do you think the happy girl lives?

Construction of Test I. Test I consisted of 12 adjectives, 10 of which were selected from *Lundholm's* list. *Lundholm's* complete list could not be used as a comparison of his adjectives with *Horn's* word list[5] showed that only 10 of the 48 were understood by children before entering the first grade. The two additional words, 'tired' and 'happy,' could have been used by *Lundholm* and in all probability would have been classed under groups III and IV respectively. Table I shows the lists of adjectives used.

The teacher introduced the experimenter, telling the children that he wanted them to take part in a little experiment or exercise. *E* then gave the following instructions:

---

[5] Horn, M. D., *op. cit.*

> If I were to ask you to draw a line, you might make one like this (*E* draws a straight line on the blackboard); others might make one like this (*E* draws a wavy line); still others might make lines like these (*E* draws a series of variously shaped lines). You see there are lots and lots of lines which can be drawn. (*E* erases *all* the lines from the blackboard.) Now I am going to give you some words and I want you to draw a line which you think means the same thing as the word. Think what the word means and then draw a line which means the same as the word. If you can't be sure, guess. Now in space No. 1 I want you to draw a *quiet* line, a line which you think is quiet. (Pause, while line is being drawn, then—) There, that's fine. Now, in the second space I want you to draw an angry line, a line which you think is angry (*etc.*).

Care was taken not to build up the empathic state through emphasis upon the word. While there were few children in any grade who did not understand the meanings of the words, there were many who confused the words. The word 'merry' was frequently used as 'marry.' Another word, 'hard' caused trouble in that the children tended to think only in one way, *i.e.*, if they thought 'hard sidewalk' they did not think 'hard lessons,' 'hard to get up,' *etc.* Accordingly the word was always used in several sentences.

TABLE I. *List of adjectives used in Test I*

1. Quiet	5. Sad	9. Hard
2. Angry	6. Dead	10. Strong
3. Lazy	7. Playful	11. Weak
4. Merry	8. Tired	12. Happy

The order was arranged by chance.

Mimeographed data sheets were provided on which were recorded age, sex, school, grade, *etc.* Test I required some limiting device such as the rectangular space on the data sheet since the size of the angles, length of curves, and degree of slant are dependent upon the size of the space in which they are drawn.

Devising a scoring system for the line drawing test presented a number of difficulties. Against the procedure of scoring responses in terms of the group response were the objections: that the 'group' would include many ages from kindergarten to college subjects and that the scores would be based to a certain extent upon the non-empathic responses of the lower grades. The best procedure seemed to be that of building a scale of values for each item based upon the work of *Lundholm*, *Poffenberger*, and *Barrows*.

A scale of seven steps was constructed varying from one conforming to the greatest extent with previous investigators' classifications to one representing the opposite feeling. Examples of the scoring scale for items 3 and 12 are shown in Fig. 2.

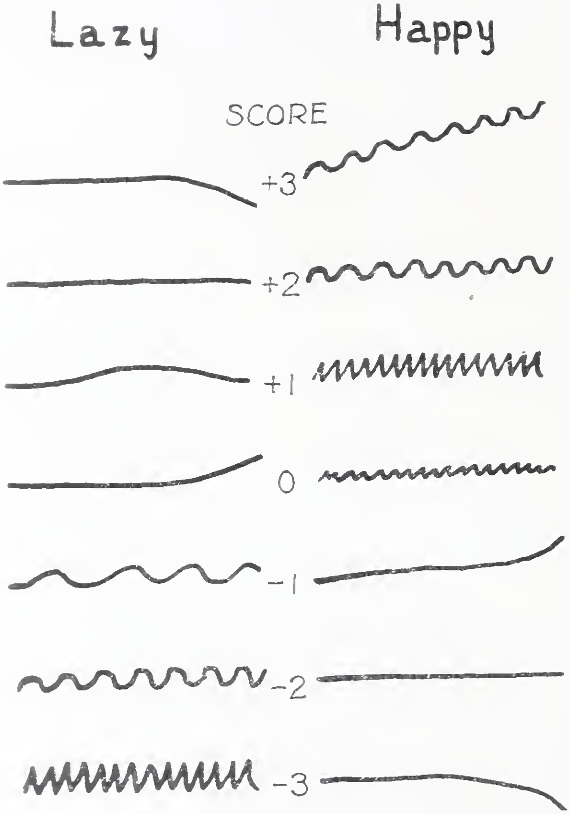


FIG. 2. Examples of scoring scale.

Construction of Test II. Test II reversed the dual choice technique employed in the preliminary tests which had proved to be too easy. Twelve lines were drawn and six colors placed upon white drawing paper 6 x 8 in. This was mounted in turn upon heavy gray cardboard so that the drawing was centered when it was projected by a Spencer delineascope on a Translux daylight screen. This projection enabled everyone in large groups to see clearly the direction, form, and type of line. In smaller groups the cards were presented directly to the subjects. There was no difference in results between these methods of presentation. Each child was provided with a mimeographed list of paired words one of which was the correct response. The instructions for giving the test were as follows:

> Last week you drew some lines for me when I gave you some words. Today I am going to show you some lines which I have drawn and I shall also give you pairs of words for each item. I want you to place a circle around the word of each pair which you think means the same thing as the word. For example, the first two words are happy-sad; if you think the line is happy, put a circle around the word happy, if you think the line is sad, put a circle around sad. If you cannot be sure, guess.

The test was scored as the number right. Line, color, and total scores were treated separately. The items and correct responses are shown in Plate I.

Construction of Test III. Test III was based on the assumption that when the subject is shown a line or color, the first words of a brief spontaneous report of his reactions would indicate the empathic response and its strength. In order to eliminate the imaginative response, an entirely different mental response from the feeling response which it is assumed appears first, the time interval for response must be extremely short.

The following instructions were used in giving the test:

> The first time I has here, I gave you some words and you drew lines meaning the same as the words; the next time I showed you some lines and colors and you put circles around the correct word. Now we are going to try something new,—I am going to show you some lines and I want you to tell me how you think you would feel if you felt the way the line feels. This may sound silly but let's try it, just *tell me how you think you would feel if you felt the way the line feels* or tell me how you think the line feels. This is a little harder than the other tests, so do the very best you can; if you can't be sure, guess.

Scoring such a test is difficult and is open to the objections that the scorer may change his criterion or that his scoring is too objective. The first difficulty was avoided by formulating a scale by which all papers would be scored. A frequency chart was made of the responses of 600 subjects. When these were studied and compared with the reports of other investigators of adult subjects, there was in general a close agreement between the highest frequencies and what has been accepted as the most

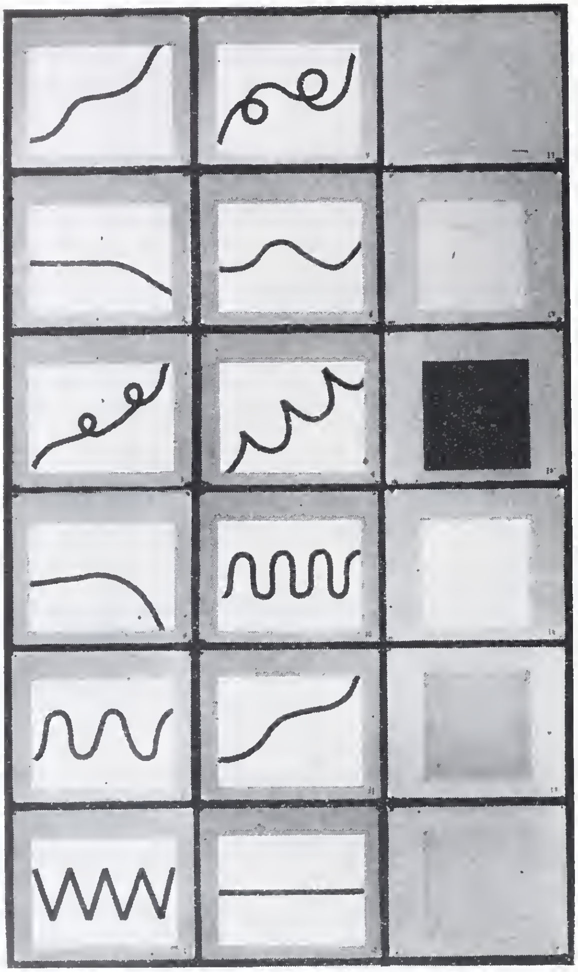


PLATE I. Items employed in Test II.

Correct answers: (see text for explanation).

1. Happy
2. Weak
3. Merry
4. Old
5. Active
6. Noisy
7. Lively
8. Not Angry
9. Playful
10. Not Lazy
11. Not Tired
12. Hard
13. Warm [orange red]
14. Weak [blue green]
15. Heavy [violet blue]
16. Happy [orange yellow]
17. Noisy [red violet]
18. Not Playful [green yellow]

common empathic reaction to lines and colors. There seemed also to be a relationship between the frequency of the words or expressions and their suitability. Thus for a line to which the majority gave the word *merry,* the word *good,* given by a smaller group, would be less appropriate, while the words *feeling bad,* given by a very small group, would be inappropriate. This relationship of frequency to suitability was not always the rule, however, for there were many *unique* responses. These might correspond to adult reactions as the response *agitated* rather than the usual one, *angry.* "The line feels like its going to have a new dress" is unique but is as correct as *joy.*

The nine lines and five colors selected are shown in Plate II.

After a complete analysis of the responses by three judges, a seven point scale was drawn up upon which each of the 75 or more responses given for each stimulus was classified. Upon this standard the performance of each subject was judged.

Construction of Test IV. Children simulate the emotion being portrayed in a picture, especially in the simpler emotional states. *Langfeld* says, "The meaning of facial expression is learned from the movement of one's own face, or a tendency toward such movement as suggested by the lines of the face observed.[6]" It has been suggested that there can be no interpretation of facial expression unless the state is induced in the subject in a kind of implicit behavior. Inasmuch as the above involves empathy and as lines and surface shapes are present in the picture of a facial expression, it was thought that such a test might be related in some way to the other tests being employed. Those tests involved single lines and no surface shapes. In the last test the interpretation would be based possibly upon the configurational pattern of the lines *and* shapes.

Ruckmick (13) prepared a set of 34 pictures expressing as many emotional states. *Gates* (6) made use of ten of the *Ruckmick* pictures[7] in testing the ability of 902 school children and 86 college students to interpret facial expression. The present investigation selected 20 of the *Ruckmick* pictures for

---

[6] Also found in H. S. Langfeld, The judgment of emotions from facial expressions, *J. Abnorm. Psychol.*, 13, 1918, 172–184.

[7] These may be obtained from the C. H. Stoelting Co., Chicago.

Test IV, the more difficult ones, those which were quite similar in expression to others in the accepted group being omitted. Care

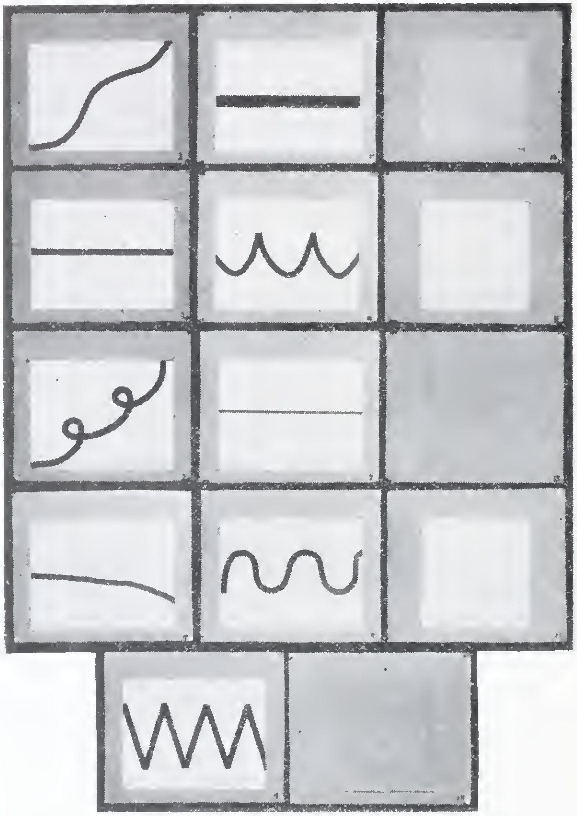


Plate II. Items employed in Test III.

Best answers: (see text for explanation).

1. Happy, glad, strong
2. Strong, hard, heavy
3. Playful, merry, happy
4. Sad, weak, lazy, tired
5. Strong, heavy, hard [*]
6. Happy, merry, playing
7. Weak, thin, light
8. Happy, merry, jolly
9. Mad, angry, hot
10. Cool, light, weak [blue green]
11. Happy, merry, gay [orange yellow]
12. Sad, unhappy, weak [green yellow]
13. Happy, clean, cool [white]
14. Warm, hot, happy [orange red]

[*] Usually prefixed by the word 'real,'—'real' strong, 'real' heavy, *etc.*

was taken to avoid sequential presentation of two similar facial expressions. The instructions were similar to those employed by *Gates*, being adapted, however, to the child-level.

The instructions for giving the test were as follows:

> Today you are to take the fourth and last test. This one is different from the others. There are 20 pictures of an actress. I want you to look at each one and tell me what you think the lady is doing, what she is thinking about or how she is feeling. Answer any one of the questions, "*What is the lady doing?*" or "*What is the lady thinking about?*" or "*How is the lady feeling?*" Write as much as you care to in the space provided,—one word or a whole sentence. If you can't be sure, guess. [In the lower grades, *i.e.*, up to and including the 3rd grade, the children were asked: "How do you think your mother would feel, or be thinking about, if she looked the way the lady looks? Now how do you think the lady is feeling, thinking about, *etc.*"]

Test IV encountered the same scoring problem as Test III. It was assumed that there is more than one correct response for each facial expression and that the responses will range themselves along a scale from good to bad. Frequencies were found for each response to each item, using the data from 500 subjects. The majority in each instance corresponded with *Ruckmick's* list. On the basis of frequency and suitability scoring scales were constructed. The word suitability is used in a special sense in that it would be incorrect to depend upon frequency alone as such a procedure would not allow for *unique* responses, such as accompany the pictures shown in Fig. 3.

Reliability of the scoring system. The responses were referred to the key or scale in Tests II, III, and IV so that the only subjective error occurring would be in the scoring of the unique responses. It was necessary in these cases to understand the general meaning and then to find the word on the list best fitting the response. For example, the statement "She looks like she has been home all day with nobody to talk to," was set down as *lonesome* and scored accordingly. The original classification was kept as objective as possible by the utilization of the frequency charts, by reference to the work of previous investigators, and by employing three trained judges in the classification of each response on the scale of values.

Test I involved greater subjectivity in judgments. Each response had to be matched to the scale of values for that item. As no two lines are exactly alike, interpretations must be made as to curvity, direction, character of irregularities, *etc.* Accordingly 100 papers were scored by two competent judges. The correlation between their judgments was .92±.006 which would be considered satisfactory.

A study was made of the comparative merits of the scoring used by *Guilford* and *Guilford* (4) in a test for students of design and the scoring system used in this investigation for Test I. Three factors were noted in their responses: (1) the form of the lines, as angle, curve, wave, or line; (2) the direction, as horizontal, up, and down; (3) the type, heavy, medium or light.

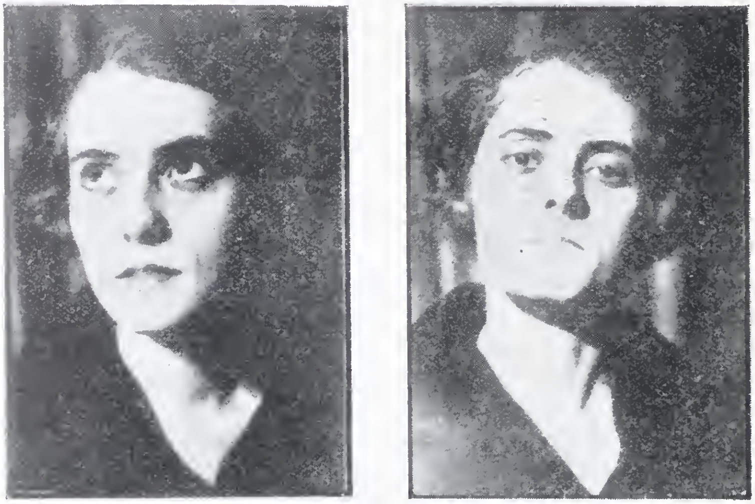


FIG. 3. Children's reactions to the above pictures involved such unique responses as: "She is thinking of God," "She is looking at her father and she loves him," "She is thinking of her sweetheart," given for *adoration;* and "She is 'high-hatting' some one," "She is acting stubborn (or saucy, smarty, haughty, snobbish, or determined)," given for *defiance.*

After a statistical study of the responses of a large number of subjects, weightings were determined for each of the ten variations possible with each adjective. Thus the word *sad* might produce a light line, extending downward. This would give values of two, four, and two, or a possible score of eight.

When a comparison was made of the lines illustrating the highest possible score and the worst possible score using the *Guilfords'* method and the *Walton* method, it was found that the two methods gave identical results. To find if differences occurred between these extremes a statistical study was made of the eight identical words in the two tests. The correlations between the two methods of scoring are shown in Table II. In general, the agreement is good.

TABLE II. *Correlation between two methods of scoring*

Test Items	No. of Subjects	Correlation	P.E.
Quiet	95	.713	.034
Angry	95	.690	.039
Lazy	94	.546	.051
Sad	94	.647	.039
Dead	95	.433	.057
Playful	95	.421	.057
Hard	95	.773	.029
Strong	93	.660	.039
Total Score	151	.696	.027

A study was made of the systems as used in scoring the tests of artistic and non-artistic groups. The *Walton* system of scoring had revealed differences between artistic and non-artistic children in the lower grades, but it did not reveal such differences between the artistic and non-artistic groups at the junior high and university levels.

TABLE III. *Comparison of two systems of scoring* *

		Artistic Group				Non-artistic Group				Chances in 100 of a true diff.
		No.	Av.	σ	S.D.	No.	Av.	σ	S.D.	
Children †	G	34	48.67	.98	5.71	34	48.82	.94	5.47	54
"	W	34	45.88	1.27	7.41	34	48.53	1.46	8.54	91
Adult ‡	G	39	52.64	.76	4.73	42	51.14	.64	4.15	93
"	W	39	51.72	.76	4.72	42	53.24	.94	6.11	89

* Items 1–7, 9–10 only could be used for this comparison.
† Lincoln Junior High Students.
‡ University Students.

Table III shows that only the *Guilford* system approximates a significant difference in favor of the artistic group. The higher score exhibited by the non-artistic children on the eight items when the *Walton* scoring was used are completely out of harmony with all other results obtained.

In spite of the superior diagnostic property of the *Guilford* system in the adult range, because of the agreement between the two systems of scoring, the fact that the majority of children in the lower grades seem to differ somewhat from adults in the kind of lines drawn and the fact that, since growth is characteristic of this ability, it is difficult to establish an age basis for weighting, the *Walton* system was used in the scoring.

Reliability of the tests. The results of repeating the tests are summarized in Tables IV and V and in Fig. 4. Repetition of the tests increased the average test score.

In individual cases lower scores usually were the result of 'intellectualizing' the response in an effort to improve test performance instead of continued dependence upon feelings.

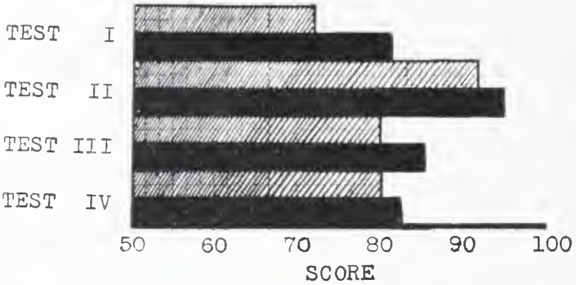


Fig. 4. Change in test score. The shaded line indicates the first trial, the solid line, the second.

Table IV. *Showing results of repeating the tests*

Test	Number	Trial 1 Av.	Trial 1 σ	Trial 1 S.D.	Trial 2 Av.	Trial 2 σ	Trial 2 S.D.
I	59	71.89	1.43	11.05	81.55	.87	6.70
II	59	16.53	.20	1.58	17.18	1.38	1.06
III	62	80.35	1.27	10.00	86.89	1.05	8.30
IV	65	80.56	.93	7.52	82.23	.94	7.45

The correlation between initial and re-test scores of 318 subjects in the Cedar Rapids schools was .40±.03. Failure to secure a higher correlation may be due to the fact that variations may occur in the reports of subjects which are due, not to the way the test is constructed or scored, but rather to the variation of the subject's feeling state as shown by *Langfeld* (11) and *Ruckmick* (13).

Curves of distribution are shown in Fig. 5, with the average and P.E. for each curve. The curve for Test II is skewed to the left, a statistical treatment of the data giving the skewness as —.55. The other curves give the following measures of skewness: Test I=.05, Test III=—.34, Test IV=.15.

TABLE V. *Reliability coefficients of correlation*

	Repeated			Split-half Method			Whole Test	
Test	N	r	P.E.	N	r	P.E.	r_x	P.E.
I	318	.40	.03	—	—	—	—	—
I	59	.26	.08	99	.32	.06	.48	.05
II	59	.43	.07	96	.31	.06	.47	.05
III	62	.31	.07	98	.59	.04	.74	.03
IV	65	.70	.04	97	.29	.06	.43	.05

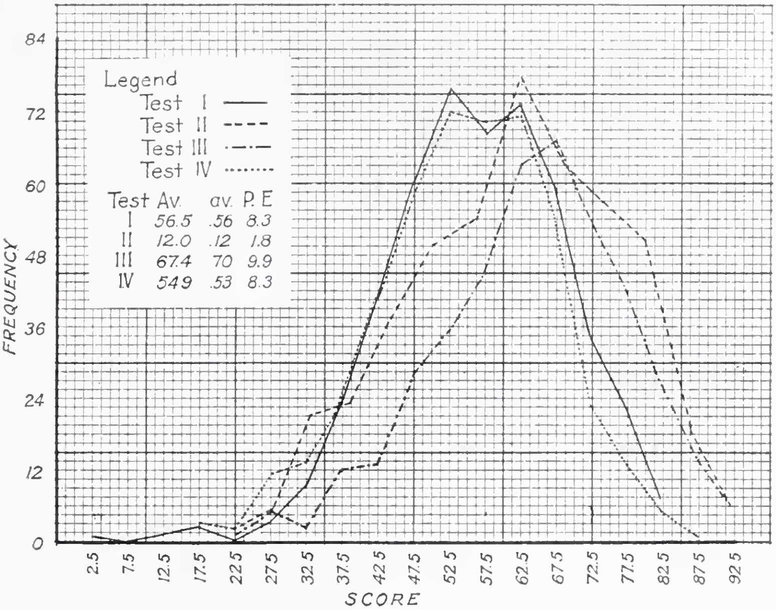


FIG. 5. Distribution of scores

III. Technique: factors investigated; appearance of the empathic response. Children in preschool laboratories of the University of Iowa and the kindergarten of Harrison School, Cedar Rapids, Iowa, were tested to discover the first appearance of the empathic response and its nature. Tests II to IV were given individually.

Growth in capacity for empathic response. The following groups were tested: grade school children from kindergarten up to VI A in Harrison School; Junior High Schools of Lincoln, Nebraska, and Des Moines, Iowa[8]; several hundred university students.

Artistic ability and empathic responses. Groups for the studies of artistic ability and empathic response were selected from four schools: first, the 'X' and 'Z' subjects[9] from Iowa City, Iowa; two groups of children, one artistic and the other non-artistic from the Cedar Rapids schools; two 'paired' groups of children representing the most and least artistic of 3,000 Lincoln school children; and fine arts and college students. This gave us an opportunity to study four artistic groups and four non-artistic groups of equal mental ability but different ages.

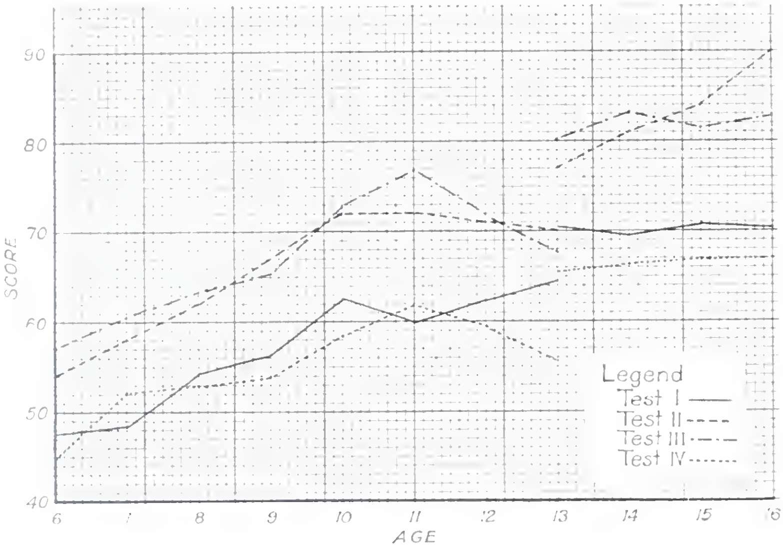


FIG. 6. Relationship between age and score.

Other factors studied. A number of our subjects had been given a color harmony test[10] during some previous experiments. The scores on this test were compared with those of Test I and

---

[8] A study made by Dr. Meier.

[9] See C. Tiebout, The psychophysical functions differentiating artistically superior from artistically inferior children, *Psychol. Monog.*, 45, 1933, No. 1, 108-133.

[10] See p. 14, footnote 13.

Part 2 of Test II to determine whether or not similar factors were involved. School grades and intelligence were compared also with empathic scores. Finally relationships between the tests and analyses of the responses to each item by all subjects and artistic and non-artistic groups were studied.

IV. Results. Fig. 6 shows the relationship between age and test score.

Genetic development of empathy. Correlations between age and empathic response are shown in Table VI.

TABLE VI. *Showing correlation with age*

Test	N	r	P.E.
I	486	.466	.025
II	473	.407	.026
III	431	.325	.029
IV	461	.416	.026

Table VII shows the average, σ and S.D. for each age-group. Those from 6 to 13 are from Harrison School, twelve to fourteen

TABLE VII. *Showing average scores at various age levels*

	Test I				Test II *			
Age-Group	N	Av.	σ av.	S.D.	N	Av.	σ av.	S.D.
6	49	47.50	1.60	8.70	47	9.68	.36	2.44
7	59	48.25	1.23	12.40	57	10.48	.31	2.35
8	83	54.15	1.23	11.25	82	11.24	.25	2.30
9	54	56.20	1.52	11.15	56	12.01	.34	2.55
10	98	62.60	1.04	10.25	91	13.02	.20	1.91
11	75	59.90	1.14	9.85	77	13.03	.24	2.10
12	41	62.25	1.38	8.85	41	12.73	.38	2.37
13	16	64.35	2.92	11.65	17	12.65	.51	2.11
12–14 †	68	68.90	1.29	10.63	66	15.23	1.80	1.47
Adult	72	77.36	1.88	9.42	66	15.96	.16	1.33
	Test III				Test IV			
6	38	57.00	2.91	17.95	48	44.90	1.81	12.50
7	38	60.40	2.59	16.00	54	52.10	1.41	10.60
8	74	63.30	1.75	15.05	78	52.85	1.11	9.85
9	52	65.30	2.01	14.50	52	53.85	1.28	9.30
10	86	72.80	1.46	13.65	83	58.25	1.14	10.35
11	75	75.85	1.38	12.95	74	61.70	1.39	11.95
12	41	71.90	2.12	13.55	40	59.35	2.07	13.10
13	17	67.80	4.04	16.65	16	55.60	2.45	9.80
12–14 †	65	81.73	1.09	8.82	66	64.23	1.32	10.74
Adult	72	82.32	1.23	10.45	68	79.57	.96	7.94

* Range on Test II is 0–18.

† Lincoln Junior High School Pupils—Compare Table VIII for results from another group.

from the Lincoln Public Schools, and the adult group from the University of Nebraska.

In a second series of experiments the study was carried into the upper age groups in the Junior High Schools of Lincoln and given to a second University of Nebraska group. These results are shown in Table VIII.

TABLE VIII. *Showing averages of junior high school* and adults for all tests*

Age-Group	Test I				Test II			
	N	Av.	σ	S.D.	N	Av.	σ	S.D.
13	53	70.35	1.46	10.69	56	13.93	.18	1.34
14	55	69.45	1.09	8.08	54	14.67	.29	2.13
15	60	70.65	1.22	9.47	61	15.03	.20	1.55
16	21	70.22	1.49	6.86	17	16.12	.25	1.02
Adults	60	74.20	1.40	10.86	59	16.15	.13	1.31
	Test III				Test IV			
13	54	80.13	1.45	10.69	50	65.48	1.77	12.48
14	56	83.13	1.41	8.91	59	66.18	1.13	8.67
15	61	81.57	1.31	10.33	56	66.75	1.55	11.57
16	20	82.80	1.56	6.96	17	66.96	2.34	9.67
Adults	63	80.99	1.96	15.53	63	78.32	1.03	8.18

* Lincoln, Nebraska, schools.

The test was also given to groups of children in the Des Moines Public Schools, Des Moines, Iowa.[11] These results are shown in Table IX.

TABLE IX. *Showing averages—Des Moines tests*

Age-Group	Test I				Test II			
	N	Av.	σ	S.D.	N	Av.	σ	S.D.
10	60	60.17	1.29	9.99	60	14.20	.20	1.56
11	36	61.83	1.48	8.82	36	13.81	.32	1.91
12	7	56.21	2.55	6.75	7	12.57	.90	2.38
	Test III				Test IV			
10	58	75.34	1.61	12.25	57	57.29	1.52	11.46
11	35	72.79	1.81	9.75	36	56.22	1.90	11.42
12	5	53.50	7.63	17.06	6	49.07	2.48	6.07

The significant correlation between age and test score supports the theory that a growth factor is involved in the empathic response. At upper levels there is in general a leveling off in age-groups, a possible exception occurring in Test II.

---

[11] Given by Dr. Meier

Constancy of response. Table IV showed that the average score was raised upon repetition of the tests. A third test, however, failed to raise the scores when the σ was taken into account.

Analysis of responses. A study of the responses revealed many individual differences. While one kindergarten child, known for his "great promise in artistic work" made scores equal to those of adults, many adult subjects made scores no better than those of children. The curves of each age-group would spread themselves over almost the entire range, being skewed to the right or left according to the age-group. There were no sex differences. The unique responses previously mentioned under scoring methods were noted. In Test III many children interpreted the lines as doing and saying things, *i.e.*, this line had fallen down and hurt itself, *etc.* It is an easy transfer from the empathic to the imaginative element,—the latter had to be used to infer the presence of the former.

One interesting feature of Tests III and IV was the dual aspect of the empathic response given by many children in the lower grades. The children's classification of the item, as pleasant or unpleasant, agreed with that of the adults, a finding in support of certain genetic theories of the emotions. The first rough classification involved such word-pairs as happy-sad, happy-sick, good-bad, or glad-sad. Older children introduced more words classifying all lines as happy-sad, mad-afraid, good-bad-sick, or glad-sad-scared-mad. To these three and four dimensional classifications other words were added so that by the fourth grade, finer differences were being made between the reactions and the rough two-dimensional classification had disappeared. There were many instances, especially in the lower grades, when a stimulus failed to arouse a response. This happened with some of our 'best' subjects, *i.e.*, subjects responding readily and giving evidence of a real empathic nature.

Test IV elicited the assumption of facial responses and bodily attitudes by many of the subjects. When pictures depicting fear or surprise were shown there was a noticeable lifting of the eyelids, an exposure of the whites of the eyes. The anger, scorn, and mistrust pictures usually resulting in a knitting of the eyebrows, a stern appearance of the face, and the slightest tensing of the muscles. A relaxation of the facial muscles followed the showing of the pleasant pictures—joyousness, adoration, interested attention, and pleasant surprise. The reactions were immediate and did not last very long.

Statistical study of responses to each item. The average score, its sigma and S.D., are given for each item of the four tests in Table X.

TABLE X. *Showing results of analysis of all items and all tests*

	Test I (N=524)			Test II (N=526)		Test III (N=475)			Test IV (N=476)		
Item	Av.	σ av.	S.D.	% Right	% Wrong	Av.	σ av.	S.D.	Av.	σ av.	S.D.
1	6.62	.12	2.75	69	31	4.40	.13	2.93	3.12	.04	.99
2	5.24	.10	2.25	68	32	4.76	.11	2.42	3.40	.06	1.41
3	4.03	.12	2.79	77	23	4.68	.13	2.90	3.17	.06	1.41
4	4.12	.10	2.34	84	16	5.51	.11	2.53	2.56	.05	1.21
5	4.21	.11	2.62	82	18	4.81	.11	2.36	2.43	.04	.92
6	4.47	.13	2.87	76	24	5.28	.13	2.85	1.66	.08	1.78
7	4.65	.10	2.24	85	15	4.60	.11	2.57	2.60	.07	1.69
8	4.31	.12	2.79	30	70	4.50	.13	2.80	4.76	.04	.83
9	4.90	.11	2.56	84	16	3.61	.13	2.75	1.66	.09	1.86
10	4.45	.11	2.59	54	46	4.33	.12	2.51	3.62	.09	1.82
11	3.98	.11	2.51	43	57	4.19	.13	2.89	2.48	.08	1.71
12	4.00	.10	2.38	79	21	5.07	.11	2.39	2.80	.07	1.60
13				84	16	4.87	.12	2.63	2.11	.07	1.54
14				58	42	5.72	.11	2.33	1.34	.06	1.32
16				31	69				3.10	.06	1.45
15				75	25				1.94	.10	2.20
17				31	69				3.81	.08	1.79
18				60	40				1.84	.07	1.55
19									3.98	.07	1.47
20									1.68	.08	1.65

Test	Range
I	0-8.3
II	0-100.0
III	0-7.2
IV	0-5.0

Items 1 and 2, lines drawn in response to the words 'quiet' and 'angry' seem too easy. Item 11 seemed low inasmuch as its direct opposite preceded it. Test II seemed much too easy for diagnostic purposes. This may have been due to the fact that it was a dual choice rather than a multiple choice situation. Test III showed high but not excessively high averages. Test IV showed six low mean scores, six which were average and eight which were high, one excessively so. This item is the picture entitled, *joyousness*, designated by most children as *happy*. It seemed to have been very commonly understood and was consequently not diagnostic. *Anger* was much less commonly understood. Some of the items in the test were too difficult.

Artistic ability and empathic response. The artist may create by using lines which he has been taught will create a certain impression or he may use lines dictated by his own feelings. Two schools of pedagogical methodology have arisen concerning the desirability of these two alternatives. Children are better subjects than adults for the acquisition of information concerning this question as direct or indirect training has had less opportunity to influence their behavior. Accordingly, the 'X' and 'Z' subjects used in previously reported experiments[12] were compared as to performance on the tests. There were only seven subjects available of the artistic group, six of which took all four tests. Of the nine non-artistic children, eight took the four tests. The results are shown in Table XI.

TABLE XI. *Comparison of X and Z subjects* *

	Artistic Group				Non-artistic Group				Chances in 100 of a true difference
Comparison	N	Av.	σ	S.D.	N	Av.	σ	S.D.	
Test I	7	65.43	2.05	5.42	9	64.00	3.17	9.52	65
Test II, Pt. 1	7	8.78	.22	.65	9	9.42	.68	1.80	76
Test II, Pt. 2	7	4.57	.45	1.18	9	4.44	.36	1.07	58
Test II	7	14.00	.95	2.51	9	13.22	.84	2.53	73
Test III, Pt. 1	6	48.00	2.70	7.64	8	45.63	3.86	9.45	69
Test III, Pt. 2	6	29.50	2.38	5.82	8	24.13	3.02	8.54	92
Test III	6	75.83	5.94	14.55	8	69.38	3.12	8.82	83
Test IV	6	57.50	5.98	14.62	8	60.63	3.24	9.16	67

* See footnote 9.

With the exception of one part of one test the differences are all in favor of the artistic children. The differences are not very significant, however, only one approaching the point of reliability. Probably the excellence of the school training which these children receive tends to overshadow certain natural tendencies and to make the responses to these tests more uniform.

Twenty-six children, considered by the art instructor to be most talented, and 20 children, considered least talented in her classes of 300 children in the Harrison School, were compared as to their ability on the tests. Intelligence test-scores were not available and testing was not permitted. Assuming, however, that general scholarship correlates with intelligence, a comparison of general scholastic averages of the two groups was made, which indicated that the two groups were equal scholastically. Differences in empathic scores were all in favor of the artistic children and, with the exception of Test III, were significant. Table XII summarizes the results of these comparisons.

---

[12] See footnote 9.

To secure children of a higher age group, art instructors of the Lincoln Junior High Schools selected the best and poorest chil-

TABLE XII. *Comparison of artistic and non-artistic children selected from 300 subjects* *

	Artistic Group				Non-artistic Group				Chances in 100 of a true difference
Comparison	N	Av.	σ	S.D.	N	Av.	σ	S.D.	
Scholarship	24	76.80	1.19	5.82	20	76.20	1.15	5.31	64.0
Test I	26	67.70	2.50	6.70	20	58.00	2.57	11.50	99.7
Test II	26	13.66	.40	2.04	20	12.80	.55	2.45	89.2
Test III	26	73.45	3.07	15.65	20	69.75	3.35	14.95	79.1
Tast IV	26	64.40	2.14	10.90	20	57.50	2.75	12.30	97.6

* Harrison School, Cedar Rapids, Iowa.

TABLE XIII. *Comparison of artistic and non-artistic groups*

		No.		Artistic Group			Non-artistic Group			Chances in 100 of a true difference
	Average on	Art.	Non-art.	Av.	σ av.	S.D.	Av.	σ av.	S.D.	
Children	Test I	34	34	67.94	1.64	9.58	69.41	1.98	11.57	65
	Pt. 1, Test II [1]	33	33	10.15	.22	1.25	9.85	.25	1.44	62
	Pt. 2, Test II [2]	33	33	5.33	.03	.18	5.09	.08	.43	100
	Comb. Test II	33	33	15.52	.23	1.33	14.97	.26	1.47	94
	Pt. 1, Test III	33	33	54.02	1.26	7.23	51.24	1.44	8.26	93
	Pt. 2, Test III [3]	33	33	29.73	.85	4.90	29.91	.74	4.23	56
	Comb. Test III	33	33	84.02	1.69	9.73	81.32	1.78	10.23	86
	Test IV	33	33	70.99	1.54	8.83	68.56	2.13	12.23	82
Adults	Test I [4]	34	72	76.09	1.07	6.70	76.81	1.11	9.40	68
	Pt. 1, Test II	36	66	11.11	.19	1.11	10.50	.15	1.20	99
	Pt. 2, Test II	36	66	5.50	.11	.68	5.46	.08	.68	62
	Comb. Test II	36	66	16.64	.23	1.37	15.96	.16	1.33	99
	Pt. 1, Test III	35	72	55.59	1.60	6.27	51.25	1.03	8.71	99
	Pt. 2, Test III	35	72	32.71	.40	2.40	30.14	.56	4.70	100
	Comb. Test III	35	72	88.22	1.21	7.19	81.60	1.23	10.45	100
	Test IV	36	68	84.33	.91	5.49	79.71	.96	7.94	100

[1] Range 0-12.
[2] Range 0-6.
[3] Range 0-41 [illegible]
[4] See Table.

dren from an artistic point of view, matching the two groups for intelligence as far as possible. Thirty-four children took Test I and thirty-three Tests II, III and IV. These results are shown in Table XIII. The differences were significant in only three instances. All but two of the eight measures, however, showed differences favoring the artistic children. It may be that by the time school children reach this age the major differences have been eliminated through their art training, perhaps indirectly by the setting up of favorable conditions in which there is use of expressions such as 'angry' line, 'happy' line, *etc.*

The tests were also administered to two adult groups, one selected by an instructor of Fine Arts, the other constituting an Arts and Science group. The results, shown also in Table XIII, show differences in favor of the artistic group, with the exception of Test I. There is also uniformity between the groups on Part 2 of Test II. In the other six instances our results show a significant difference in favor of the artistic group.

All of the items used in the tests were analyzed further to learn which were more diagnostic of artistic ability. Tables XIV–XVII show the results. These tables also reveal those items which were diagnostic in the lower grades but not in the upper,

TABLE XIV. *Showing results of an analysis of all items, Test I, for the good in art and the poor in art*

	Grade School Children [1]						Junior High School [2]					
	Good in Art [3]			Poor in Art [4]			Good in Art [5]			Poor in Art [6]		
Item	Av.[7]	σ av.	S.D.	Av.	σ av.	S.D.	Av.	σ av.	S.D.	Av.	σ av.	S.D.
1	8.15	.07	.37	6.62	.57	2.53	6.71	.42	2.42	7.60	.28	1.70
2	5.96	.24	1.21	5.46	.49	2.17	5.75	.47	2.77	6.22	.40	2.38
3	5.11	.46	2.33	4.20	.54	2.40	5.80	.25	1.43	5.38	.36	2.17
4	5.05	.37	1.94	5.53	.49	2.19	3.99	.23	1.36	4.58	.17	1.04
5	4.84	.53	2.70	4.25	.51	2.27	5.35	.37	2.16	5.82	.42	2.50
6	4.56	.49	2.51	5.46	.59	2.62	5.99	.27	1.52	5.98	.27	1.60
7	6.02	.29	1.50	5.60	.26	1.15	4.58	.29	1.69	5.45	.31	1.88
8	5.18	.54	2.73	5.46	.62	2.76	6.17	.33	1.90	5.32	.35	2.10
9	5.75	.47	2.37	4.20	.58	2.58	6.33	.34	1.98	6.75	.30	1.78
10	5.32	.41	2.10	3.78	.37	1.66	4.48	.40	2.32	6.02	.42	2.52
11	5.96	.33	1.70	4.34	.53	2.38	5.88	.29	1.69	6.13	.22	1.33
12	5.70	.41	2.09	3.92	.53	2.39	5.36	.32	1.86	5.05	.31	1.87

[1] Harrison School, Cedar Rapids, Iowa. Ages 8–11.

[2] Junior High Schools, Lincoln, Nebraska. Ages 12–15.

[3–6] 26, 20, 34, 36 children respectively.

[7] Range: 0–8.3.

TABLE XV. *Showing results of an analysis of all items, Test II, for the good in art and the poor in art*

	Grade School Children[1]				Junior High School[2]			
	Good in Art[3]		Poor in Art[4]		Good in Art[5]		Poor in Art[6]	
Item	% right	% wrong	% right	% wrong	% right	% wrong	% right	% wrong
1	54	46	80	20	94	6	75	25
2	88	12	60	40	88	12	78	22
3	96	4	75	25	100	0	95	5
4	96	4	85	15	94	6	95	5
5	100	0	80	20	97	3	100	0
6	81	19	90	10	94	6	97	3
7	96	4	95	5	100	0	95	5
8	12	88	25	75	53	47	42	58
9	96	4	95	5	91	9	97	3
10	73	27	70	30	75	25	72	28
11	62	38	35	65	34	66	42	58
12	96	4	90	10	97	3	97	3
13	100	0	100	0	100	0	92	8
14	77	23	80	20	97	3	92	8
15	31	69	10	90	94	6	100	0
16	85	15	70	30	91	9	78	22
17	54	46	30	70	50	50	53	47
18	77	23	80	20	100	0	100	0

[1-6] See footnotes, Table XIV.

TABLE XVI. *Showing results of an analysis of all items, Test III, for the good in art and the poor in art*

	Grade School Children[1]						Junior High School[2]					
	Good in Art[3]			Poor in Art[4]			Good in Art[5]			Poor in Art[6]		
Item	Av.[7]	σ av.	S.D.	Av.	σ av.	S.D.	Av.	σ av.	S.D.	Av.	σ av.	S.D.
1	4.52	.53	2.71	4.38	.68	3.04	4.98	.43	2.51	5.82	.38	2.20
2	4.89	.48	2.49	5.28	.50	2.26	6.04	.32	1.88	4.66	.38	2.27
3	6.09	.42	2.18	5.52	.62	2.69	6.65	.18	1.07	6.95	.38	.22
4	6.65	.33	1.67	5.64	.55	2.26	6.33	.32	1.87	6.67	.22	1.28
5	5.17	.45	2.30	5.10	.51	2.31	5.78	.31	1.76	5.82	.43	2.54
6	5.77	.59	3.05	4.68	.68	3.06	5.53	.36	2.07	5.75	.39	2.27
7	5.77	.50	2.56	5.40	.54	2.42	6.73	.15	.88	5.44	.43	2.51
8	4.85	.56	2.85	5.64	.54	2.43	5.27	.19	1.10	5.89	.46	2.70
9	4.61	.40	2.04	3.72	.66	2.97	5.49	.35	2.00	4.62	.44	2.57
10	4.85	.57	2.59	4.98	.73	3.27	4.98	.53	3.04	5.54	.43	2.50
11	4.43	.59	3.05	5.40	.68	3.12	5.38	.32	1.84	5.58	.28	1.65
12	5.12	.45	2.30	5.64	.56	2.52	5.89	.35	1.99	5.08	.39	2.27
13	5.26	.41	2.11	4.32	.60	2.67	6.69	.23	1.33	6.56	.23	1.35
14	6.00	.33	1.70	5.94	.48	2.16	7.02	.10	.60	7.02	.15	.72

[1-6] See footnotes, Table XIV.

[7] Range 0–7.2

TABLE XVII. *Showing results of an analysis of all items, Test IV, for the good in art and the poor in art*

	Grade School Children[1]						Junior High School[2]					
	Good in Art[3]			Poor in Art[4]			Good in Art[5]			Poor in Art[6]		
Item	Av.[7]	σ av.	S.D.	Av.	σ av.	S.D.	Av.	σ av.	S.D.	Av.	σ av.	S.D.
1	3.57	.18	.96	3.38	.17	.79	3.63	.15	.90	3.91	.10	.60
2	2.98	.31	1.60	3.30	.24	1.09	3.99	.16	.97	4.00	.18	1.07
3	4.15	.21	1.10	3.66	.25	1.14	3.73	.21	1.24	3.59	.21	1.28
4	3.20	.16	.84	2.58	.23	1.06	2.99	.17	1.00	2.95	.15	.93
5	2.43	.16	.82	2.42	.19	.87	2.93	.14	.83	2.72	.08	.51
6	2.90	.34	1.74	1.94	.40	1.82	2.97	.34	1.79	2.14	.32	1.80
7	3.02	.30	1.55	2.90	.38	1.70	4.25	.21	1.22	3.88	.40	1.48
8	4.88	.06	.32	4.40	.27	1.24	4.98	.01	.11	4.99	.01	.08
9	2.70	.32	1.64	2.80	.42	1.92	2.18	.27	1.60	2.34	.30	1.76
10	4.38	.20	1.04	4.24	.30	1.36	4.88	.05	.29	4.87	.01	.11
11	3.37	.19	.98	2.53	.48	2.15	3.83	.29	1.68	3.63	.25	1.48
12	3.92	.31	1.59	3.86	.32	1.45	3.97	.20	1.19	3.74	.21	1.25
13	2.87	.41	2.12	1.82	.29	1.34	3.08	.30	1.74	2.59	.30	1.82
14	1.56	.30	1.56	1.16	.24	1.11	2.43	.31	1.78	2.56	.24	1.47
15	2.90	.31	1.58	3.30	.26	1.20	3.63	.19	1.11	3.84	.13	.77
16	2.90	.36	1.85	2.18	.52	2.34	2.41	.35	2.04	3.20	.31	1.87
17	3.80	.34	1.75	3.88	.36	1.53	4.71	.17	.96	3.99	.27	1.62
18	2.58	.31	1.60	1.86	.39	1.76	3.33	.27	1.53	2.82	.28	1.65
19	4.35	.22	1.13	4.52	.19	.89	4.05	.19	1.14	4.06	.22	1.29
20	2.50	.33	1.70	2.04	.35	1.57	2.98	.27	1.55	2.55	.30	1.77

[1–6] See footnotes, Table XIV.
[7] Range: 0–5.

those diagnostic in the upper grades but not in the lower, and those diagnostic in both. Items 4, 6, and 8 of Test I, items 1, 7–10, 12–16, and 18 of Test II (in fact the whole test), items 2, 4, 8, 10–12 of Test III, and items 2, 5, 9, 15, 17, and 19 of Test IV could be eliminated from the tests and all of the remaining items combined into a single test.

Relationships between tests. To determine if the tests actually succeeded in measuring the same thing, coefficients of correlation

TABLE XVIII. *Correlations between various tests*

Comparison of	N	r	P.E.
Test I–II	480	.328	.028
Test I–III	440	.315	.028
Test I–IV	461	.344	.027
Test II–III	612	.532	.020
Test II–IV	638	.688	.016
Test III–IV	613	.516	.020
Test I–Pt. 1–III	439	.302	.029

between the various tests were computed. These relationships are presented in Table XVIII.

Test I seems to bear less of a relationship to the other three tests than they do to each other. In general the correlations are sufficiently high to warrant the assumption that to a fairly high degree they are measuring the same thing.

Relationship to other factors. Since a larger number of the children of Harrison school had been given a color harmony test,[13] the correlation was found between this test and Test I. The result, shown in Table XIX, indicates no relationship whatever. To determine if the color reactions were based on a sense of fitness rather than on a feeling for the color, the color harmony scores were correlated with the results of Part 2 of Test II yielding the result shown in Table XIX. Inasmuch as there was little relationship between single color preferences and sensitivity to color harmonies,[14] it would seem that the various color experiences are quite distinct.

TABLE XIX. *Showing correlations with scores on Color Harmony Test*

Comparison of	N	r	P.E.
C H—Test I	194	.163	.046
C H—Pt. 2, Test II	223	.058	.045

Correlations between grades made in school and Test I and between grades in art and Test I indicated no significant relationship. These are presented in Table XX. There were no significant relationships to sex and to intelligence.

Genetic development of empathy. In addition to the statistical

TABLE XX. *Showing correlations between grades in school and Test I*

Comparison of	N	r	P.E.
Test I—Grade in Art (lower grades) (I B–IV B incl.)	225	.094	.045
Test I—Grade in Art (upper grades) (IV A–VI A incl.)	213	.093	.047
Test I—All Grades (upper and lower)	436	.098	.037

---

[13] Walton, W. Sensitivity of children and adults to color harmony, *Psychol. Monogr.*, 45, 1933, No. 1: 51–62.

[14] *Ibid.* p. 60

results of the investigation presented heretofore, observation of the performance of children of many ages in the test situation led the author to suggest the following theory of the genetic development of empathy.

Results from the preliminary tests given to pre-school children suggest that very young children *feel themselves into their entire environment.* It may be that, beginning soon after birth, children develop a process of abstracting themselves from their environment, feeling themselves less and less into their surroundings. The empathic tests tap these more primitive methods of reacting which, although they undoubtedly play some part in adult daily behavior, are more subtle in their influence. It is a common observation that very young children are 'bubbling over' with ways in which the aeroplane, automobile, train or tractor 'go.' Voice, arms, legs, and indeed the whole body are used to describe one of these vehicles. Even the mention of the name is sufficient to start these physical processes. In adults abstraction has taken place so that they no longer feel themselves into the situation to such an extent, but in all probability the name of an object will arouse incipient movements. To a lesser extent this is true of adjectives.

Children do not, however, have very well established communities of ideas. This would account for the low scores of very young children. To the pronounced word *angry* the child grasps his pencil, frowns, tenses his muscles and vigorously makes a *wavy* line. The empathic response is present but may be poorly expressed. However, his responses more and more resemble those of other children, and as we proceed up the age scale we gradually approach adult performance. In all probability a great many factors are operative to determine this growth,—certainly age is not the only factor. Accounting for the growth of the response on the basis of the acquisition of a community of ideas is probably not adequate unless it be assumed that children's experiences are the same.

V. Conclusions. A battery of four tests was constructed, designed to measure the empathic response in children and adults. Test I involved the drawing of lines in response to adjectives.

Test II was a dual choice situation in which lines and colors were shown the subjects and they were required to encircle the correct word. Test III called for a spontaneous response to lines and colors, the response involving a kind of introspective analysis of the subject's reactions toward lines and colors. Test IV employed 20 of the *Ruckmick* pictures of facial expression and followed the *Gates* technique for the instructions. The results obtained by giving the tests to kindergarten, grade school, and junior high school children, college students, groups of artistic and non-artistic subjects, lead us to the following conclusions:

1. The empathic response begins at a very early age and may manifest itself in the reactions of kindergarten children.

2. Growth is characteristic of this behavior and is gradual in nature indicating a certain constancy.

3. Individual differences appear as early as kindergarten age and are manifest in all age-groups. They concern themselves for the most part with the ability to make the response.

4. Unique responses, *i.e.*, those differing radically from the typical and in many instances probably involving a great use of the imagination, an accompaniment of the response, are found in all tests (except Test II) and at all age-levels.

5. Young children show a tendency to think in terms of pleasantness-unpleasantness, utilizing a dual type of response. Gradually they add other dimensions until a large range of responses is acquired.

6. Bodily reactions accompany the empathic response, as revealed by facial expressions and bodily tensions.

7. In general there seems to be some relationship between artistic ability and the four tests; exceptions are present in certain groups.

8. There is a definite relationship between the varions tests. The tests are fairly reliable, but not as high as is desirable.

9. Grades in school, intelligence, grades in art, reactions to color harmony tests and sex do not seem to be related to the empathic response.

10. The results are suggestive of further work and introduce

a possible diagnostic factor in the study of artistic and non-artistic children.

Bibliography

1. Allesch, G. J. von Die aesthetische Erscheinungsweise der Farbe. *Psychol. Forschung*, 6, 1925, 1–91, 215–281.
2. Birkhof, G. D. Æsthetic measure. Cambridge, Mass., Harvard Univ. Press, 1933, pp. 6–14.
3. Bullough, E. On the apparent heaviness of colors. *Brit. J. Psychol.,* 1, 1907, 111–152.
4. Chandler, A. R. Beauty and human nature. N. Y., D. Appleton-Century, 1934.
5. Cox, G. J. Art for amateurs and students. N. Y., Doubleday Page, 1926, p. 21.
6. Guilford, J. P., and Guilford, R. B. A prognostic test for students in design. *J. Appl. Psychol.,* 15, 1931, 335–345.
7. Gundlach, C., and Macoubrey, C. The effect of color on apparent size. *Amer. J. Psychol.,* 43, 1931, 109–111.
8. Gates, G. S. A test for ability to interpret facial expression. *Psychol. Bull.,* 22, 1925, 120.
9. Hilpert, R. S. A critical study of a modified free association test given to students in the Department of Art Education. (Unpubl. Univ. of Minn. study.)
10. Koch, B. C. The apparent weight of colors. N. Y., Methodist Book, 1928.
11. Mogensen, M. F., and English, H. B. The apparent warmth of colors. *Amer. J. Psychol.,* 37, 1926, 427–428.
12. Lundholm, H. The affective tone of lines: experimental researches. *Psychol. Rev.,* 28, 1921, 43–60.
13. Langfeld, H. S. The æsthetic attitude. N. Y., Harcourt Brace, 1920.
14. Poffenberger, A. T., and Barrows, B. E. The feeling value of lines. *J. Appl. Psychol.,* 8, 1924, 187–205.
15. Ruckmick, C. A. A preliminary study of the emotions. *Psychol. Monog.,* 30, 1922, No. 3, 30–35.
16. Walton, W. E., and Morrison, B. M. A preliminary study of the affective values of colored lights. *J. Appl. Psychol.,* 15, 1931, p. 301.
17. Warden, C. J., and Flynn, E. L. The effect of color on apparent size and weight. *Amer. J. Psychol.,* 37, 1926, 398–401.

THE VALIDITY AND RELIABILITY OF A MODIFIED FORM OF THE McADORY ART TEST FOR USE AT LOWER GRADE LEVELS [1]

by

Mildred Dow Voss

Introduction. An investigation of the development of æsthetic appreciation carried on in the second, third, fourth, and fifth grades, required an art judgment test which could be applied satisfactorily to these grade levels. The *McAdory* Art Test [2] is standardized for public-school ages and grades above Grade III. Preliminary trial indicated that the subject-matter of this test seemed suitable for lower grades. Accordingly an attempt was made to determine if the standardized form or a modified form of this test might be suitable for use in the second and third grades.

Sampling. The subjects used in this study represented the entire enrollment of the specified school grades of two Iowa towns; Villisca, a town of about 2,000 population, and Grinnell, a town of 5,000 population. The second-, third-, and fourth-grade results are from Villisca, the fifth-grade results from Grinnell.[3] Both towns are primarily dependent upon the surrounding rural population for existence, having few manufacturing industries. Grinnell represents a greater socio-economic range because of the presence of a small college. It has also a greater amount of public-school art education.

---

[1] This study was made concomitantly with the writer's other investigation reported in this volume and is an indirect phase of the research program, Genetic Studies in Artistic Capacity. Since it is a by-product of another investigation, the data are more limited than they would have been had the primary object been the standardization of the test at these levels.

[2] McAdory, M. Art test. N. Y., Bur. of Publ., Teach. Coll., Columbia Univ., 1929.

[3] Acknowledgment is made to Superintendents H. E. Dow and C. E. Humphrey and to the principals and teachers of the Villisca and Grinnell Public Schools for their co-operation in this study.

Test program. In Grades II, III, and IV, four tests were given: (1) an introductory test on the basis of which two groups were formed, one of which received art training, the other of which did not; (2) a follow-up test to both groups, after which the second group received a different type of training; (3) a second retest four weeks later; (4) a third retest following the summer vacation, approximately a year later than the original test. In Grade III, three tests were given: (1) an introductory test as a basis for division of groups; (2) a retest after a training period of three weeks for one group; (3) a second retest after a training period of three weeks for the second group.

Table I presents the number of children in each grade, the number of art tests administered, the mean CA (in months), and the mean intelligence and MA scores (in months) derived from the *Kuhlmann-Anderson* Group Intelligence Test (4th ed.).

Table I. *Description of grade groups*

Grade	No. of children	No. of art tests administered	Mean IQ	Mean MA	Mean CA
II	30	103	102.5	91.1	90.1
III	40	154	100.4	102.1	103.5
IV	33	123	100.7	114.9	114.5
V	85	252	102.5	130.0	128.5

Grade V as normative. Grade V represented the highest grade level in this investigation and is a year above the lowest grade on which the test is standardized. It was assumed, because of the *McAdory* standardization, that the test is reliable and valid at this level. The measures of reliability and validity obtained from the administration of the test at the fifth-grade level were used as a norm. It was proposed to test whether the standardized form of the test continued to give like measures of reliability and validity as it was applied to progressively lower levels, or, if not, what modifications could be adopted to insure a comparable validity and reliability.

Administration. *McAdory* allows considerable variation in administration of the test. The following procedure was adopted in Grade V as a method which retained most of the advantages of individual testing while also utilizing the time-saving advantage of the group test.

The children were seated in rows in their customary classroom manner. Each child was furnished with the standard record blank and was given an equal number of the test plates, the number being determined by the number of plates divided by the number of children in the group to be tested. Care was taken that the plates were in consecutive order, *i.e.*, the first child had Plates 1 to 7, the second child, Plates 8 to 14, etc., and that this order was retained throughout the test period. The following directions were given:

> Find the record sheet on your desk like this (holding up record sheet). Put your name on the line after 'Name.' Put .... (number of the grade) on the line after 'Grade,' and .... (name of school) on the line after 'school.' (The date was stamped in by the experimenter and the birth date was secured from school records.)
>
> I have given you some cards. Each card has four pictures on it. Do you see that each picture is different? In this test you are to look at each picture very carefully and decide which picture you like best, which one you like second best, and which one you like third best.
>
> Let me show you how to do it with this card (hold up Plate 71). See how each picture is different (point out differences). At the top of this card is the word 'Plate' and a number. Find this number on your record sheet (hold up record sheet). It is in a little square and under it are four lines. The first line begins with the word 'Best.' Suppose we say that we like B best (point out on plate). We put the letter B on this first line under number 71. Everybody do this. (Cover up B on Plate 71): Now which of these three pictures do you like best? Let us say that D is best. Put a D on the line beginning with the words 'Next Best' on your record sheet. (Cover up D): Now which of these two pictures do you like better? Let us say that A is better. Put A on the line beginning with 'Third Best' on your record sheet. Now only C is left so that we put the letter C on the line beginning with 'Worst' on the the record sheet. You see the picture we liked best is on the line named 'Best,' the picture we like next best is on the line named 'Next Best,' the picture we like third best is on the line named 'Third Best,' and the picture that we did not choose is on the line named 'Worst.' Now let me see if your paper is correct. (Check each child's paper.)
>
> Here is Plate 36. Let us say that we like D best, C next best, A third best, and B worst. (Write on blackboard.) Put this answer where it should go on your record sheet. (Check each child.)
>
> Now take the clip off your cards and lay it on your desk. Leave it there until I tell you to use it again.
>
> Each of you has different plates. Look at the top of each plate and find the number; then find this number on your record sheet. Put your finger on this number so that you will not lose it. Look at your pictures on the first plate. See how each one is different; then choose the picture you like best. Put the letter of this picture on the first line. Choose the picture you like next best. Put the letter of this picture on the second line. Choose the picture you like

third best. Put the letter of this picture on the third line. Then put the letter of the picture that is left on the last line. Then go on to the next picture. When you have finished all of your pictures, turn them over on your desk so that the white side is up.

(After each child has finished): Leave your record sheets lying on your desks and pick up your plates. Now all of you have finished your pictures. We will have to change so that each person has new ones. Listen carefully and do just as I tell you. Each person in this row, pass your pictures to the one back of you. The person at the back of the row brings his pictures to the front. Now he gives his pictures to the first person in the *second* row. Now, everyone in the second row, pass your pictures to the one back of you. The last one in the second row brings his pictures to the first one in the *third* row. Everyone in the third row passes his pictures back. The last one brings his pictures to the first one in the *fourth* row. Now see if this row can do it. This row. Now I will take the pictures of the last one in the last row and bring them over to the first person in the first row. Now you have new pictures and can start again.

(After each child has again finished): Let us see if you can change pictures correctly. First row, change pictures. Second row. Third row. Fourth row. Fifth row.

(After each child has again finished): Change plates.

(At conclusion of class period): We will have to stop today. Put your record sheet on top of your pictures and clip them together with the clip on your desk. Now wait until I have completely finished telling you what to do. I want you to pass your papers forward up to the first of the row. Wait. As you take the papers from behind, put your paper on top. You (to the last child in row) pass your paper forward. Now you (second child from back of row) put your paper on top and pass them both forward. You (third child) put your paper on top of both. This is very important or the papers will be out of order for us tomorrow. Have your paper *on top* before you pass the papers on. Now, second row, pass your papers. Third row. Fourth. (Collect papers being careful to keep them in order.) Tomorrow we will go on from where we stopped today.

(Testing time was broken into three half-hour periods on successive days. Care was taken not to hurry the children.)

Adequacy of the test at the fifth-year level. *McAdory* gives norms for each age and grade based on scores made by New York City children.[4] The mean score of grade V-B girls was 103.3; the *McAdory* norm for grade V-B girls is 106. The mean score of the fifth-grade boys was 93.9; the corresponding *McAdory* norm for grade V-B boys being 102. To determine if a discrepancy of age accounted for these lower scores, the *McAdory* norm for grade V-B girls of a mean CA, corresponding to the mean age of the Grinnell girls, of 10 yrs. 6 mos. was determined as 114.0; and the *McAdory* norm for grade V-B boys of a mean CA, corresponding to the mean age of Grinnell boys, of 10 yrs. 9 mos. was determined as 105.0. The Grinnell scores compare even more unfavorably with the norms when age is considered. However, it must be remembered that the environmental discrepancy is undoubtedly great.

---

[4] Siceloff, M. M., Woodyard, E., *et al.* Validity and standardization of the McAdory Art Test. N. Y., Bur. of Publ., Teach. Coll., Columbia Univ., 1930, pp. 10–16.

The *McAdory* Art Test is divisible into two comparable forms; Plates 1 to 36 comprising Form A, Plates 36 through 72 comprising Form B. The reliability of the test for Grade V was determined by correlating the score on Form A with the score on Form B. The resulting coefficient of .71±.02 became, on application of the Spearman-Brown prophecy formula, a reliability for the whole test of .83. *Carroll* and *Eurich* report a reliability coefficient, obtained by similar means, of .82 for 100 college students.[5]

As one group received no training between the first and second test, a second measure of reliability was determined by correlating each subject's score on the initial test with his score on the retest three weeks later. For 41 children, *r* was .87±.03.

McAdory[6] quotes correlations between initial and retests on groups receiving instruction in the interim as measures of reliability. The correlations were as follows: 101 art school students retested after one week, .73±.04; 95 art school students retested after two months, .63±.05; 54 women college students retested after a week, .73±.05; 47 women college students retested after a year, .59±.07. A similar coefficient was obtained for the subjects in this investigation by correlating the first and second test of the group receiving instruction. The resulting coefficient of .64±.06 for 44 subjects compares favorably with the reliabilities quoted above.

Modification of administration. From the foregoing, it would seem safe to conclude that the *McAdory* Art Test in its standardized form is as reliable at the fifth-grade level as in the higher age-levels. The problem now was to determine if the standardized form could be applied without loss of reliability to lower grade levels or if modifications of administrative technique were necessary. Two modifications seemed possible to increase the reliability of the test should it fail to be applicable in the standardized form: (1) administration as an individual rather than as a group test; (2) diminution of the difficulty of separate items by eliminating the more difficult intermediate choices and requiring the child to choose only between the best and worst items.

---

[5] Carroll, H. A., and Eurich, A. C. Abstract intelligence and art appreciation. *J. Educ. Psychol.*, 23, 1932, p. 216.

[6] Siceloff, M. M., Woodyard, E., *et al., op. cit.*, p. 27.

To determine the effect of individual testing as opposed to group testing, a modification of procedure was followed in the administration of the test in the fourth grade. On the first 22 plates of the test the child was tested individually, on the remaining 46 plates he was tested in a group as in Grade V. Following is a description of the modified procedure on the first 22 plates:

(There was first a brief preliminary exercise from the *Lewerenz* Tests in Fundamental Abilities of Visual Art, Test 1, Recognition of Proportion, Items 1 and 2.) I have some pictures I want to show you today. Here is one of them. There are four cups and each cup is different. You are to choose the cup that you think is the best looking, the one that you would like to have most. Which of these four would you choose? Now suppose I cover this cup (the one just chosen). That leaves these three. If you could have just one of these cups that are left, which would you choose? (Cover up one chosen): Now which of these two cups that are left would you choose?

Now do you see what we have done? We have picked the one you liked the very best (pointing), then the one you liked second best (pointing), the one you liked third best (pointing), and that left the one you liked least of all (pointing). We are going to do that with all of these pictures. It is a kind of a game. Haven't you often picked out pictures you liked best in magazines or the kind of cars you liked best? That is what we are going to do here.

Now let's try another one. Which of the bowls do you like best? Second best? Third best? Look at each one carefully.

(The *McAdory* plates were then presented with the following introduction accompanying each. The answers were recorded by the experimenter).

Plate 1: Here are four sets of letters. If you look carefully you will see that each set of letters is printed differently. Which of the four sets is printed the best? (Cover up one chosen): Now which of these three sets of letters is printed best? (As before): Now which of these two sets of letters is printed the better?

Plate 2: Here are four tables. Suppose you were buying a new table for your living room at home. Which table would you choose as the very best looking table? But suppose you could not have that table, which of these three tables would you choose as the prettiest? Suppose you had only these two to choose from, which of these would you choose as the prettier table, the one you would like to have?

Plate 3: Suppose your mother were to tell you that you might have a new sweater and the clerk showed you these four sweaters. Which one would you like to have? But suppose that this sweater had already been sold and you must choose from these three. Which one would you take then? But suppose

that your mother said that sweater was too expensive and that you must take one of these two. Which one would you choose?

Plate 4: Now suppose we were going to put a fence around the school house yard. Look at these four fences and tell me which one of them would be the very best looking fence. But if we couldn't get that fence, which one would you choose next? Suppose that the only two fences we could get were these two that are left. Which one of these would you choose as the better looking fence?

Plate 5: Now suppose you were going to buy a new car. There were these four kinds that you could buy. Which would you choose first? Which would you choose second? Which third? Which fourth?

Plate 6: Suppose you were going to buy a new dress (if a girl) or suppose your sister was going to get a new dress (if a boy). You had these four dresses to choose from. Which dress would you like to have most? Then suppose that your mother said you might have another dress. Which one of these three dresses would you choose then? Then suppose that your mother said that you needed just one more dress and no more. Which dress would you choose as the prettier dress of these two?

Plate 7: Suppose your father is going to buy a new store building down town. There are these four buildings that he likes. Which one of the four do you like best? If he isn't able to buy that one, which of the four would you tell him to choose next? Then which of the two that are left would be the better looking? So you will tell your father to buy this one first, if he can, because it is the best looking; this one next because it is the next best looking; this one third because it is the third best looking; and not to buy this one at all because it isn't at all good looking.

Plate 8: Now suppose that your mother is having a birthday soon, and you want to buy a very pretty plate for her birthday present. Which one of the four plates would you choose as the very prettiest? Which one as the second prettiest? Which one as the third prettiest?

Plate 9: Suppose that your little baby brother is going to get a new suit. Which suit will you choose as the best looking for him? Which suit as the second best looking? As the third best looking?

Plate 10: These are pictures for the wall. You want to get a new picture to hang on the wall of your schoolroom. Which will you choose as the best looking picture? Which as the second best looking picture? Which as the third?

Plate 11: Suppose that you are getting a pitcher for your mother. Which will you choose first? Which second? Which third?

Plate 12: (Omitted.)

Plate 13: Here are some more little boys' suits. Which one will you choose first for your little brother? Second? Third?

Plate 14: These are books. This part is the front of the book and this is the part of the book that bends when you open it. You see that they are colored differently. Which cover would you rather have on your books? Which is the next prettiest book cover? Which is the prettier of these two?

Plate 15: Here are some more plates. These are much like those we saw before only this time they are colored. Which is the prettiest plate of them all? Which is the prettiest of these three? Which of these two?

Plate 16: Suppose your mother is buying a new dress. Which one of these four dresses would you like her to have? Which of these three dresses? Which of these two dresses?

Plate 17: This is a picture of a Dutch lady. You see each picture is colored differently. In this picture her skirt is almost white, in this one it is darker,

and in this one it is very dark. You see that the walls are colored differently; the hanging on the wall is different, and her cap is different in each picture. Which of these four ways of coloring makes the best looking picture? Which of these three ways of coloring makes the best looking picture? Which of these two?

Plate 18: Here are four bowls, all alike except that they are colored differently. Which do you think is the very best looking bowl, the one you would like to have most? Which is the second best looking bowl? The third best looking bowl?

Plate 19: Here are some more dresses. Which of these dresses would you like your mother to have? Which of these three dresses would be prettiest for her? Which of these two dresses?

Plate 20: Here are some pictures of people. You see that they did not put in the faces so that we have to decide which is the best looking picture from the rest of the picture. Which one do you like best of all? Which one do you like best of these three? Which of these two?

Plate 21: Here are some rugs. Which rug would you like to have best for your room at home? Suppose you had only these three rugs, which would you like best? Which of these two?

Plate 22: Which of these four davenports would you like for your living room at home? Which of these three davenports? Which of these two?

Plate 23: Do you see that the trees and bushes are planted around this house in different ways? Which way do you think they are planted around the house best? Which way in these three pictures? Which way are they planted around the house better in these two pictures?

Reliability of the standardized test in Grade IV. The first problem was to determine if the standardized form of administration of the test was as reliable in Grade IV as in Grade V. The reliability coefficient between the score on Plates 24 to 47 inclusive and the Plates 47 to 72 inclusive was computed for each child in Grade V and in Grade IV. The assumption was that the comparative reliability coefficients of these exercises would reflect the comparative reliability of the entire test administered in such a manner. For Grade V the reliability coefficient was .64±.02. Applying the Spearman-Brown prophecy formula for triple length, the reliability for the whole test was .84. For Grade IV the corresponding coefficient was .65±.04 which becomes .85 for the whole test. It would appear that the application of the test in its standardized form is as reliable in Grade IV as in Grade V.

Effect of individual administration of the test on reliability. To determine if individual administration of the test increased its reliability, the scores on the first 23 plates in Grade IV were correlated for the first and second tests, the second and third, and the third and fourth. The resulting correlation was .60±.05, which is raised to .82 for the whole test by the Spearman-Brown prophecy formula. To find the reliability of a similar series of exercises administered in the standardized manner, each child's score was determined for Plates 37 to 60 inclusive, omitting Plate 48, of Form B of the test. This gave a duplicate series of approximately equivalent difficulty. The resulting correlation was .70±.04, which becomes .88 for the whole test. It would appear that administration as an individual test, at least in this grade, diminishes the reliability of the test rather than increases it.

Effect of elimination of intermediate stage of difficulties on reliability. In Grades II and III the children were tested as in Grade IV, individually on the standardized form of the test on the first 22 plates. To determine if elimination of the intermediate degrees of difficulty on the plates improved the reliability of the test, the first 23 plates were scored for best and worst items only, *i.e.*, if the best item was placed before the worst in the rating, the child received a score of 1, if the worst item was placed before the best, the child received a score of 0. Thus the possible range of scores was 0 to 68. These scores were also correlated for the first and second, the second and third, and the third and fourth tests for each grade. The results are presented in Table II.

TABLE II. *Comparison of the four-item form and the two-item form of the McAdory Art Test in Grades II, III, and IV* [7]

Grade	No. of cases	*r* for four-item form, 22 items	*r* for whole test	*r* for two-item form, 22 items	*r* for whole test
IV	89	.60±.05	.82	.61±.05	.82
III	112	.44±.05	.70	.49±.05	.74
II	73	.42±.07	.68	.47±.07	.73

Either form of the test is of equal reliability in Grade IV, but in Grades III and II the two-item form of the test has greater reliability.

---

[7] These coefficients of reliability are lowered due to the fact that scores were correlated on tests between which a deliberate effort was made to change the child's score. However, as the same influences operated on both forms, they may be used to determine the comparative reliability of the two forms.

Effect of individual administration on the two-item form. In order to determine if individual administration of the test increased the reliability of the two-item form of the test in Grade IV, the two-item scores on the first 22 plates were compared with the two-item scores determined for each child on Plates 37 to 60 inclusive, omitting 48, on Form B in Grade IV. The reliability coefficient on repeated tests on the two-item form administered individually (Plates 1–23) was .61±.05; on the duplicate form administered as a group test (Plates 37–60), .64±.05. This result is similar to the previous finding that individual administration of the test appears less reliable than group administration.

Half-half reliability of four-item and two-item forms. Correlation of scores on repeated tests shows that individual administration of the test fails to improve reliability but that omission of intermediate difficulties does improve reliability. To check further on the comparative reliability of these two forms, the coefficient of correlation between Form A and Form B of the test was computed for both the four-item and the two-item scores. The reliability for the four-item form for Grade IV is .68±.03 which is a reliability quotient of .85 for the whole test. Again the two-item form demonstrates a slight superiority in reliability.

Administration of Plates 24 to 72 inclusive in Grades II and III. The test was administered as an individual test in the second and third grades. The first 22 plates were administered as in Grade IV. Plates 24 to 72 inclusive, however, were administered in the two-item form. Each plate was covered by a white cardboard with openings which permitted only the best and worst item of each plate to be seen.[8] The child indicated simply the one he preferred of the two. Accordingly, no comparison of the reliability of the individual and group form of administration of the test could be made in these grades.

Reliability of modified forms of the test in lower grades compared with the reliability of the standard form in Grade V. On the basis of the foregoing results the test form giving maximum reliability would seem to be the group two-item form in Grade IV, and the two-item form in Grades II and III. As explained above, relative reliability of the group or individual form in these latter grades was not tested. The reliability for the whole test in each of these grades was determined by correlating the scores on Form A and Form B of the test and applying the Spearman-Brown prophecy formula. These reliabilities are compared with that obtained for Grade V in Table III.

---

[8] In this, as in all other uses of the term, 'best' refers to best as determined by *McAdory*.

TABLE III. *Relative reliability of the standardized form of the test for Grade V and the modified forms of the test for Grades II, III, IV*

Grade	No. of tests	Form	Half-half reliability	Reliability whole test
V	250	Group McAdory (4)	.71±.02	.83
IV	123	Group McAdory (2)	.68±.03	.81
III	153	Individual McAdory (2)	.68±.03	.75
II	96	Individual McAdory, (2)	.75±.03	.86

The reliability of the test used in these modified forms in the lower grades seems to be approximately equivalent to that obtained by using the standardized form of the test in Grade V, and apparently the test may be legitimately used as an extension to the lower levels as far as reliability is concerned.

The reliability coefficients obtained from repetition of the test are shown in Table IV. Many of the reliability coefficients quoted by *McAdory* and other investigators are based upon the repeated test performance of groups having received instruction in the interim. The accompanying investigation offered an opportunity to test the effect of intervening training upon these

TABLE IV. *Reliability coefficients for repetition of McAdory Art Test for groups with and without training*

Experiment	Weeks intervening between tests	*r* between Test I and II of group receiving training		*r* between Test I and II of group not receiving training
1—McA^2	20	.64±.06		.77±.04
II McA^2	3	.87±.03		.79±.04
Retest	20		Group E	.81±.04
			Group C	.69±.06
III—McA^1	3	.72±.05		.87±.03
IV McA^1	3	.74±.05		.86±.03

coefficients. Accordingly reliability coefficients were computed for trained and untrained groups for each experiment.

Training decreases the reliability coefficients on the test in all of the experimental situations except Experiment II. Either group of coefficients, however, is sufficiently high to compare favorably with the coefficients obtained by *McAdory* as previously cited.

Reliability of the sub-tests of the McAdory Art Test. The *McAdory* Art Test may be subdivided into three sections testing ability to recognize comparative merit in form, in notan, and in color. As results were computed for each of these sections separately, it was desired to find some indication of the reliability of the score on each section of the test. Table V presents the results.

TABLE V. *Reliability of the sections on Form, Notan, and Color of the McAdory Art Test as measured by the correlation between Form A and Form B*

			Correlation between Form A and B			Raised for Whole Test		
Grade	Form used	No. of tests	Form	Notan	Color	Form	Notan	Color
V	McA (4) Group	250	.65±.03	.27±.03	.28±.04	.79	.64	.44
IV	McA (2) Group	123	.77±.03	.32±.06	.31±.05	.87	.49	.47
III	McA (2) Indiv.	153	.78±.02	.25±.05	.29±.06	.88	.37	.45
II	McA (2) Indiv.	96	.74±.03	.28±.06	.12±.07	.85	.44	.21

McAdory[9] using as subjects two art-school groups, R and S, 101 and 95 in number respectively, and 59 persons in Hunter College, Group T, found the following reliability coefficients between the 12 plates of each section in Form A and the 12 plates of each section in Form B:

	Group R	Group S	Group T
Form	.611	.577	.341
Notan	.257	.238	.235
Color	.421	.362	.378

It is rather surprising that, as a whole, the correlations obtained in these lower grades are higher for the sub-tests than the correlations obtained by *McAdory,* although the reliability correlations for notan and color tend to be rather low. The total number of exercises involved, however, is not large so that high reliability would not be expected.

---

[9] *Ibid.,* p. 28.

The reliability coefficients obtained from repetitions of the sub-tests are shown in Table VI. Reliability coefficients were computed for both trained and untrained groups for each experiment.

TABLE VI. *Reliability coefficients for repetition of the sub-tests of the McAdory Art Test for groups with and without intervening training*

Experiment	Weeks intervening between tests	*r* between Test I and II of group receiving training	*r* between Test I and II of group not receiving training
		Form	
I—McA (2)	20	.63±.06	.73±.05
II—McA (2)	3	.83±.04	.76±.04
Retest	20		Group E .73±.06
			Group C .83±.03
III—McA (4)	3	.63±.06	.73±.05
IV—McA (4)	3	.77±.06	.79±.04
		Notan	
I—McA (2)	20	.73±.05	.59±.07
II—McA (2)	3	.67±.06	.49±.08
Retest	20		Group E .50±.09
			Group C .65±.06
III—McA (4)	3	.73±.05	.59±.07
IV—McA (4)	3	.60±.07	.66±.06
		Color	
I—McA (2)	20	.35±.09	.41±.09
II—McA (2)	3	.63±.07	.44±.08
Retest	20		Group E .70±.05
			Group C .35±.09
III—McA (4)	3	.35±.09	.41±.09
IV—McA (4)	3	.47±.09	.52±.08

It can be seen that training decreased the reliability coefficients on the sub-tests, Form and Color, except in Experiment II. The reversed situation is true on the sub-test, Notan, where training increases the reliability coefficient except in Experiment IV. The difference in this latter, however, is small.

Conclusions regarding reliability. From the foregoing results we may conclude that the standardized form of the test is (1) as reliable in Grade V as in upper grades and ages; (2) as reliable in Grade IV as in Grade V. Individual administration decreases rather than increases the reliability of the test on both the four-item and the two-item form. On the first 22 plates the reliability obtained by correlating tests one and two of each test series on the four-item and on the two-item forms is approximately equal. On the entire test, however, the reliability obtained by correlating Form A with Form B is slightly greater on the two-item form. In Grades II and III the two-item form of the test has greater reliability as determined by correlation of the scores of the first 22 plates on the first and second tests of each series. This two-item form has approximately the same reliability as the standard form in Grade V.

Validity of the McAdory Art Test for lower grade levels. The validity of the application of a modified form of the *McAdory* Art Test to second-, third-, and fourth-grade levels is more difficult to establish. Teachers' ratings or art grades seemed unlikely criteria as these would be based upon ability to draw rather than to appreciate art. In Grade V the children were allowed 15 min. to sketch with show-card colors. The painting scores obtained from the resulting productions were ranked in nine degrees of merit on the basis of good composition by ten competent judges. These rankings correlated with the art appreciation score on the *McAdory* Art Test, .34±.06.

The following criteria were adopted as being indicative of the validity of the modified form of the test: (1) a comparatively high correlation of the two-item form with the four-item form; (2) a comparatively high correlation of the *McAdory* Art Test scores with the total of the scores on three tests of æsthetic judgment; (3) an increase in percentage of successes within successive grades; (4) similar correlations with intelligence, MA, and CA on the two-item and four-item forms of the test.

Correlation of the two-item and four-item forms. To determine how closely the two-item and four-item scores approximated one another, the test was graded according to both methods, and correlations between these two sets of scores were computed for each grade. These are shown in Table VII.

The correlation between the two forms of the test is high and approximately constant from grade to grade.

TABLE VII. *Correlation between the two-item form and the four-item form of the McAdory Art Test for each grade*

Grade	No. of tests	*r* for whole test	No. of tests	*r* for items 1–24	Probable *r* for whole test
V	248	.88±.09			
IV	124	.93±.08	126	.74±.03	.90
III			154	.71±.03	.88
II			103	.78±.03	.92

Correlation with the total score of a battery of tests. In addition to the *McAdory* Art Test, two short additional tests were administered to the children in the second, third, and fourth grades. These included Part I, Test 1 on Recognition of Proportion of the *Lewerenz* Tests in Fundamental Abilities of Visual Art,[10] and 12 paired items of the *Whorley* Unity Scales.[11] In adapting the *Whorley* scales for the purpose of this investigation, item 1 was paired with item 7, 2 with 6, 3 with 5 for each form of the two scales, thus forming 12 pairs of test items.

These tests were combined into a total score. As the length of each varied, the tests were given approximately equal weight by equalizing the variability according to a method described by *Garrett.*[12] As the respective variabilities were *McAdory*, 8.88, *Whorley*, 1.84, *Lewerenz*, 1.99, the measures on the latter two tests were multiplied by 5. The weighted scores of the three tests were then totaled for each child and the score on each test correlated with the total score. The resulting coefficients for 365 measures were as follows:

Score on all three tests with	McAdory Art Test	.79±.01
	Lewerenz Test	.77±.01
	Whorley Test	.73±.02

As all three tests purported to measure the same general function, the fairly high agreement of the *McAdory* Art Test with the total score seems to indicate that the test is valid to a fairly satisfactory degree.

---

[10] Lewerenz, A. S. Tests in fundamental abilities of visual art. Los Angeles, Research Service, 1927.

[11] Whorley, K. An experimental investigation of the sensitivity of the pre-school child to compositional unity. *Psychol. Monog.*, 45, 1933, No. 1, Plates I and II.

[12] Garrett, H. E. Statistics in psychology and education. N. Y., Longmans Green, 1925, p. 279.

Decreasing difficulty of test items. A third test of validity was secured by the determination of relative difficulty of test items from grade to grade. In addition to the second, third, and fourth grade groups, previously described, the test scores of 35 junior and senior high school students, and 34 adults were included in this portion of the study for comparative purposes. The percentage of correct responses of the two-item form of the test was secured for each plate for each grade. These percentages were converted into P.E. distances above or below the mean on a normal curve.

Instead of giving the P.E. value for each item for each grade, Table VIII presents a general survey of the number of test items passed by more than 50 per cent of the distribution for that grade, the number of items passed by less than 50 per cent, and the average P.E. value of the test items for that grade.

TABLE VIII. *A comparison of the difficulty of test items in successive grade groups*

Grade	No. of items passed by more than 50 per cent of the distribution	No. of items passed by less than 50 per cent of the distribution	Av. P.E. value of all items
II	34	34	.03
III	38	30	—.10
IV	40	28	—.21
V	57	21	—.47
Junior and senior high school	67	1	—2.42
Adults	68	0	—3.15

Table VIII shows that there is progressive decrease in the difficulty of the test items from grade to grade. Such a decrease would seem to indicate that the test is validly extended downward for use in the lower grades.

Correlation of art score with intelligence and age factors. A fourth test of the validity of the modified form of the test was to determine if the correlations between the modified test scores and IQ, MA, and CA in each grade resembled the correlations obtained using the standardized form of the test in the fifth grade. The correlations for each grade are presented in Table IX.

TABLE IX. *Correlation of art score with other measures in each grade*

Grade	Test Form	IQ	MA	CA
V	Four-item	.29±.06	.38±.06	.19±.06
IV	Four-item	.60±.07	.56±.07	—.12±.10
IV	Two-item	.31±.10	.28±.10	.22±.11
III	Two-item	.20±.10	.24±.09	.08±.10
II	Two-item	.29±.11	.19±.11	.18±.11

The correlations obtained using the standardized form of the test in Grade IV indicate that intellectual factors are involved to such an extent that the test is no longer a valid measure of æsthetic judgment. The two-item form presents more satisfactory relationships, which in Grades II and III seem sufficiently similar to those obtained in Grade V to warrant the conclusion that the modified form of the test does not involve other factors which destroy its validity.

Conclusions regarding validity. The results of the examination of the validity of the test add additional justification for the use of the two-item form in the fourth grade in order to eliminate the high correlation with the intelligence factor. The other results seem, in general, to justify the conclusion that extension of the *McAdory* Art Test to lower grade levels in a modified form has not impaired the validity of the test.

THE MEASUREMENT OF QUALITY IN CHILDREN'S PAINTING BY THE SCALE METHOD[1]

by

CAROLYN TIEBOUT

I. Introduction. While tests of artistic appreciation have been devised satisfying the scientific requirement involving control of all variables under consideration, less objectivity has seemed possible in the measurement of artistic ability or the ability to produce. An approach to problems involving the artistic ability of children, however, made it incumbent upon certain of the investigators (8, 9, 11), whose studies are reported in this monograph, to develop, in so far as possible, a reasonably reliable means for determining status in artistic production. For this purpose the scale of values method was chosen as offering the most objective and most practical treatment available. While this method can not eliminate the necessity for a certain degree of subjectivity in judgment, it furnishes the judges with a more or less definite set of values on which to make decisions of merit. The superiority of this method over the direct judging technique has been demonstrated by *Brooks* (2). Using as instruments the drawing scales of *Kline-Carey* (5) and *Thorndike* (10), *Brooks* found the inaccuracy of rating to be reduced by one-half of what it was when ratings were made without the use of a scale. A scale of values offers a further advantage in that, once established, it may be used for other groups of similar material at different times and by different persons, while ratings without a scale, no matter how expert the judges, are limited to the particular data used.

---

[1] This study represents a co-operative program preliminary to the investigations of *Saunders*, *Meier* and *Stinson*, and *Tiebout*. Invaluable contributions in constructing the scales were made by Dr. Norman C. Meier, the director of the study, and by the group of research associates and assistants in the Psychology of Art department.

The whole project forms a part of the Genetic Studies in Artistic Capacity sponsored by the Carnegie Foundation for the Advancement of Teaching and financed by the Carnegie Corporation.

Of the available published drawing scales, namely, the *McCarty* (7), the *Thorndike* (10), and the *Kline-Carey* (5 and 6), the latter is the most recent and advanced and is the only scale which includes a section on composition in which scoring is based on the achievement of æsthetic qualities. The *McCarty* scale also includes compositions, but these are considered, not as æsthetic representations, but as indications of ability to express ideas vividly and accurately, and are so judged.

II. General plan for scales. While the composition feature of the *Kline-Carey* scale was fitted to the purpose of this study, in which the child's ability to achieve artistic effects was to be measured, the fact that the ten samples representing the work of Grades I to VII were arranged in a single continuum did not make possible as fine a discrimination as was desired between the products of a single grade, particularly at the lower age-levels. It was therefore the plan of this study to devise a new scale, or rather a series of scales, in which the paintings for each of the grades from one through seven, were treated separately. The color element absent in the *Kline-Carey* samples was also added since it was deemed an important one in achieving artistic quality. Selecting the painting of compositions and devising a specific scale for this activity perforce limits the measure of artistic ability to this type of performance. It was the opinion of the investigators, however, that free illustration would offer, for a single performance, more opportunity for creative expression and the realization of achievement of artistic possibilities than other activities such as the representation of a single object, the designing of a border, and similar exercises.

Following the *Kline-Carey* plan, certain stories were selected as a basis for imaginative composition. This was done to secure sufficient uniformity in the paintings, to facilitate their judgment, and to insure the inclusion of compositional qualities. Story lists of from six to twelve titles were selected for each grade separately after consultation with the teacher and principal to select those most familiar to the majority of the children in the respective grades. Each child was permitted to select from his grade list the stories which he wished to illustrate or, if he preferred, he could illustrate incidents from a single story. The titles chosen were as follows:

Grade I—The Little Red Hen, Old Mother Hubbard, Peter Rabbit, The Three Bears, Billy Goats Gruff, Little Miss Muffett.

Grade II—The Little Red Hen, Old Mother Hubbard, Peter Rabbit, The Three Bears, Billy Goats Gruff, Little Red Riding Hood.

Grade III—Jack and the Beanstalk, Old Mother Hubbard, Peter Rabbit, The Three Bears, Billy Goats Gruff, Hiawatha.

Grade IV—Jack and the Beanstalk, King Arthur Stories, Old Mother Hubbard, Peter Rabbit, Robin Hood, The Three Bears.

Grade V—The Trojan War, The Pied Piper of Hamelin, King Arthur Stories, (Rip Van Winkle), William Tell, Robin Hood, Paul Revere's Ride, Winter Sports, Hannibal Crossing the Alps, Old Mother Hubbard, (Hansel and Gretel), (The Golden Fleece). Also may be substituted Billy Goats Gruff, Little Red Riding Hood, Little Red Hen.

Grade VI—The Trojan War, The Pied Piper of Hamelin, King Arthur Stories, Rip Van Winkle, Robinson Crusoe, Gulliver's Travels, William Tell, Robin Hood, Paul Revere's Ride, Winter Sports, Old Mother Hubbard, (Hansel and Gretel). Also Billy Goats Gruff, Little Red Riding Hood, Little Red Hen (one of these substituted for Hansel and Gretel).

Grade VII—The Trojan War, The Pied Piper of Hamelin, King Arthur Stories, Rip Van Winkle, Robinson Crusoe, Gulliver's Travels, William Tell, Robin Hood.

III. Subjects. To secure a sufficiently large and adequate sampling of the art products of children in the grades chosen for study, it was decided to use a group of approximately 100 subjects for each grade, normally distributed as to intelligence, with an average IQ approximating 100. With the assistance of the Division of Research and Pupil Adjustment, an elementary school, judged as best meeting these requirements, was chosen from the public school system of Des Moines, Iowa. With the exception of Grade I, however, the number of subjects for whom complete data were obtained in Phillips, the school selected, was from 20 to 30 short of that desired and the IQ average was from two to four points below 100.[2] In Grades II through VI subjects were therefore added from another school, the general IQ level of which was estimated to be somewhat above average. Two supplementary schools in the same city were used: Hubbell, in Grades II to IV, and Perkins, in Grades V and VI. As in the case of the Phillips School, no selection of subjects was made in the two additional schools, all pupils in the single class used in each grade being included.[3] In the case of Grade VII, which is included in the Junior High School system, the desired number of subjects and IQ average were secured in a single school, Woodrow Wilson, which serves the same area as the Phillips School [4].

---

[2] Classes for those retarded in school progress were not included.

The following table which gives data concerning the entire group used will show that for the most part the conditions stipulated at the beginning of the program were met.

TABLE 1. *Data on subjects*

		Sex		CA		IQ	
Grade	No.	M	F	Av.	σ	Av.	σ
I	100	67	33	6–8	5	100.38	10.53
II	106	46	60	7–9	6	98.96	8.10
III	103	51	52	8–9	7	98.78	9.99
IV	100	49	51	9–10	7	99.82	11.28
V	100	48	52	10–9	9	99.03	10.59
VI	100	57	43	11–10	8	98.16	10.56
VII	100	54	46	12–8	8	99.39	9.90

As will be noted from the table there are an approximately equal number of boys and girls in each grade with the exception of the first where the sampling gave twice as many boys as girls. There is approximately one year's difference in CA average from grade to grade. Intelligence tests were administered and paintings collected from all grades at approximately the same time so that this constant interval in chronological age and school development would be kept between each grade.

---

[3] This class in each instance was the last half of a division of the grade as distinguished from the first half, or the B division.

[4] Grateful acknowledgment is due to Miss Estelle Hayden, Director of Art of the Des Moines Public Schools, who made the program in these schools possible, and to Miss J. Ella Baker, principal of the Phillips School, for co-operation in the testing program conducted in this school from which the majority of the subjects were drawn. Acknowledgments are also made to the principals of other schools in which subjects were secured: Miss Laura P. Matthews of Hubbell, Miss Georgia Quigley of Perkins, and Mr. Everett Davis of Woodrow Wilson Junior High School, and to the art teachers and classroom teachers in these four schools who supervised the collection of the paintings.

Simultaneous collection of data in each grade also made it possible to eliminate the time variable which might otherwise have operated to affect results.

The intelligence test used was the *Kuhlmann-Anderson* Group Intelligence Test (4th ed.), which was applicable to the seven grades and thus gave a uniform testing instrument. Further data on the procedure and results of the intelligence testing are given in the following article (**11**). Since the IQ average for each grade approximates 100 the mental age averages would approximate those for chronological age given in the table.

IV. Collection of paintings. Three paintings were made by each of the 709 subjects in the study, with the exception of the subjects in Grade I-B who made two paintings and two crayon drawings. The crayon drawings were included in view of the first graders' lack of experience with the medium of paint. Three paintings instead of one were collected from each child since it was desired not only to secure a large sampling of work from each grade but to judge later the artistic status of the individual children whose products were used in the establishment of the scales. Complete instructions were given the class-room teachers in regard to control of the painting situation.[5] Paintings were made independently, no child being permitted to see the work of another or to receive aid from the teacher. Only one painting was made at a time and no time limit was set since it was felt that a more accurate indication of a child's ability could be obtained when he had ample time to perform and where a sampling of his work was not limited to a single period, though not extended over such a period as to introduce a possible growth factor. Materials used in each grade were as follows: Tempera paints in red, blue, green, yellow, violet, orange, black, white, and brown; paper, 16″x 22″ white book stock; and ¾ in. brushes. Each painting was identified by means of a code number which indicated the grade and name of the child painting it.

V. Construction of scale. The first step in the process of scale-construction involved a preliminary sorting of the 2,227 paintings thus collected so as to facilitate the procedure for the experts who were to make the final judgments for the scale. Work on the scales covered a period of two years, the first being devoted to the scales for Grades I through IV and the second to those for Grades V–VII inclusive. While the same procedure in developing the scales was followed in both instances, the personnel of the groups of judges necessarily varied. Preliminary judgments were made by four research assistants in the division of the Psychology of Art who, in addition to being trained in scientific methods, had had one to several years' experience with children's art. The further qualification of art training was possessed by three of the judges, two in the first group working with Grades I–IV inclusive and one in the second group.

---

[5] These technical aspects were supervised by Research Associate Saunders.

These judges, who will be designated as departmental judges, were instructed to arrange the paintings for each grade in five categories of merit from inferior to superior on the basis of æsthetic quality with, among other attainments and qualities, due consideration of æsthetically significant balance, rhythm, color harmony, *etc.* Six paintings for which there was complete or nearly complete agreement in the ratings of the four judges were then selected as samples of each of the five steps of merit: inferior, below average, average, above average, and superior. In other words, there were chosen for each grade thirty typical paintings, divided equally among the five degrees of artistic quality according to the agreement of the departmental judges.

These samples, representative of the range of artistic quality from definitely inferior to definitely superior in the products of each grade, were then submitted to a group of judges, designated as experts, on the basis of whose ratings the scale itself was constructed. Fourteen judges rated the paintings for Grades I–IV inclusive and 17 judges, those for Grades V, VI and VII. Ratings by a relatively small and highly selected group of judges were considered a more reliable means for the establishment of artistic value than ratings by a larger group of less experienced judges. Those selected were qualified by their experience as art educators, with definite training in and contact with the field of children's art. Care was taken not to restrict the selection to any one school of thought in art or to limit it geographically, so that no constant influence would operate in favor of any particular bias or viewpoint in the consideration of artistic merit.

These 14 judges for the earlier grades included three professors of art education, three directors of art in metropolitan public schools, four supervisors of art in a large city system, one assistant professor of art education, two instructors of children's classes in art, one of whom is also a museum director, and one graduate student in vocational guidance and major in art in a mid-western university. Four of the group are also listed in the *Biographical Directory of American Artists* (1).

While the 17 judges for the later grades do not represent a group of as high a rank in the field of art education, they are definitely qualified as judges by their art backgrounds and their extensive experience with children's art work. One of the number is a director of art; five are art supervisors; seven, teachers of art; and four, graduate assistants with art training and teaching experience.

These judges were instructed to arrange the 30 paintings in each grade in five ranks of artistic merit from inferior to superior. In judging the paintings the experts were advised to give due consideration, among other qualities, to the attainment of rhythm, balance, unity, and other æsthetic qualities,[6] rather than to technique or realistic representation. The final decision of each judge, according to instructions, was to be based on his judgment of each composition as a whole, using those criteria of artistic merit with which he was most familiar and in which he placed most confidence. Since the judges were regarded as qualified experts in the field a more definite description of the criteria for arranging the paintings was not deemed necessary or advisable. Certain mechanical aspects of the judging process were clearly demonstrated so as to facilitate the procedure for the judge. Complete instructions were as follows:

> To insure uniformity of conditions for judging the children's paintings, it would be much appreciated if the following suggestions were kept in mind.
>
> There are 30 paintings for the fifth grade in folder marked Grade V; 30 also for each of Grades VI and VII.

---

[6] Referring to the attainment of these qualities in æsthetically significant, rather than in obvious, stereotyped or dully formal ways.

> Take the paintings of one grade at a time and arrange enough space on a large table or two desks for five groups. Then distribute the paintings into five groups in accordance with artistic merit from left to right as
>
> | Group 1 | Group 2 | Group 3 | Group 4 | Group 5 |
> |---|---|---|---|---|
> | Poor | Definitely inferior to average | Average | Definitely superior to average | Superior |
>
> It is suggested that the attainment of rhythm, balance, unity, or other æsthetic qualities be given due consideration rather than technique or realistic representation. The final decision should represent the rater's best judgment of each composition as a whole, using those criteria of artistic merit with which he is most familiar and in which he places most confidence.
>
> It might be found helpful to examine the entire 30 before beginning the sorting so as to secure some idea of their general nature. After sorting a greater degree of accuracy is usually obtained if each group is re-examined in turn.
>
> When each pile is sorted finally, then go through each and arrange the paintings within the pile in order of merit. Then tabulate. The first number (stamped on back) will be the best in the group, the second, the second best, *etc.* Please record legibly and accurately.[7]
>
> When all grades have been completed, shuffle all the pictures for Grade V together so that no trace of your ranking remains; then do the same for Grades VI and VII.

In order to obtain an average of the degree of merit ascribed to each painting by the experts, arbitrary values of 1 to 5 were assigned to the five ranks or groups, one representing the inferior and five the superior. For the scale it was desired to secure paintings most nearly representative of each quality step from 1 to 5, hence those paintings were selected the average for which equaled or closely approximated these exact numerical values. In its final form the scale was amplified to nine steps by the inclusion of paintings at the mid-point values, *i.e.*, paintings for which the average fell at or near 1.5, 2.5, 3.5 and 4.5. At certain points on the scales more than one painting averaged the required value. These duplications of example were retained as further aid in judgment. The total number of paintings included in the scales was 11 in Grade IV, 13 in Grades I, III and V, 14 in Grades II and VI and 16 in Grade VII. In five cases no painting attained the necessary average value; but since these fell at the extremes of the scales, four at the upper and one at the lower, extrapolation was not difficult.[8]

---

[7] Data secured from the order of merit ranking of the judges were not used since practical considerations made it necessary to abandon the use of the more elaborate statistical treatment of the data by the probable error method employed by Thorndike and others.

[8] Failure to secure samples for the scale at the highest value resulted from a

The reliability of the judgments of the experts in rating the 120 sample paintings for Grades I–IV inclusive is indicated by a correlation of .94±.01 obtained between the average ratings of the first half of the group of judges and the average ratings of the second half, while for the 90 sample paintings in Grades V–VII inclusive the correlation secured in the same manner was .93±.01. Using the Spearman-Brown prophecy formula (3, p. 269) to determine the reliability of ratings for the entire group of judges in both instances, correlations of .97 and .96, respectively, were obtained. The average of the ratings of the 120 sample paintings by the departmental judges correlated with the average of the expert judges' ratings .93±.01 and for the 90 samples, .90±.01. These correlations are interpreted as indicating a high reliability for the method employed in the scale construction which was based entirely on the averages of individual judgments. In addition, they lend confirmatory evidence as to qualifications of the departmental judges, since these judges agreed almost as closely with the average rating of the experts and these later agreed among themselves. The correlations between departmental and expert judges, however, are doubtless higher than would result from a correlation of ratings of the entire group of over 2,000 paintings, since only those paintings were selected as samples for which the departmental judges agreed exactly or approximately.

VI. Use of scales. Since we desired to know the art status of the subjects whose work was used in devising the scales, their paintings were again judged by the four departmental judges who for this rating employed the scales which had been established. The average of the values assigned by the four judges was taken as the score for each painting and for the final score of each individual an average was taken of the values thus assigned to his three paintings or, in the case of Grade I, to his two paintings and two crayon drawings. For the final judgments of the paintings using the scales, correlation between the

---

tendency on the part of the expert judges, particularly in the grades above the second, to place a larger number of paintings in the lower than in the upper categories of merit. This was due probably to a tendency to judge the paintings according to standards achieved by children who had more specific training in the handling of tempera paint.

ratings of two judges against two judges yielded a coefficient of .79±.01 in Grades I through IV and of .73±.01 in Grades V–VII inclusive. These coefficients of correlation were raised to .88 and .84, respectively, when the reliability for the four judges was determined by applying the Spearman-Brown prophecy formula.

Data on the actual results of the applications of the scales are included in reports of the studies in which various units of the original group of subjects were used. For their investigations *Saunders* (9) and *Tiebout* and *Meier* (11) included the entire group of 100 subjects in each grade, the former using Grades I–IV inclusive and the latter the complete grade range from I–VII inclusive. In the *Meier* and *Stinson* study (8) certain selected subjects from Grades III, V, and VII were used. With the resulting consistency in the ratings of artistic ability and the testing of the same group of subjects for other related factors, more complete and comparable results were possible. The use of the scales for subjects other than those of the original group, as in the study of *Saunders,* afforded a further advantage in the attainment of a uniform definition of status in artistic production throughout the program.

Bibliography

1. AMERICAN ART ANNUAL (Ed. McGlauflin, A. C.). Washington, D. C., American Federation of Arts, 1931.
2. BROOKS, F. D. The relative accuracy of ratings assigned with and without use of drawing scales, *School & Soc.,* 27, 1928, 518–520.
3. GARRETT, H. E. Statistics in psychology and education. N. Y., Longmans Green, 1926.
4. KINTER, M. The measurement of artistic abilities. N. Y., The Psychological Corporation, 1933.
5. KLINE, L. W., and CAREY, G. L. A measuring scale for freehand drawing, *Johns Hopkins Stud. in Educ.,* 5, Part I, 1922.
6. KLINE, L. W., and CAREY, G. L. A measuring scale for freehand drawing, *Johns Hopkins Stud. in Educ.,* 5, Part II, 1933.
7. McCARTY, S. A. Children's drawings. Baltimore, Williams & Wilkins, 1924.
8. MEIER, N. C., and STINSON, R. E. Æsthetic sensitivity and productivity in art (In preparation. To be published at subsequent date).
9. SAUNDERS, A. W. The stability of artistic aptitude at the child level (this volume, pp. 126–154).
10. THORNDIKE, E. L. The measurement of achievement in drawing, *Teachers College Record,* 14, 1913, No. 5.
11. TIEBOUT, C. E., and MEIER, N. C. Artistic ability and general intelligence (this volume, pp. 95–125).

ARTISTIC ABILITY AND GENERAL INTELLIGENCE[1]

by

CAROLYN TIEBOUT AND NORMAN C. MEIER[2]

I. Introduction. In analyses of the relationship between various abilities and general intelligence, certain ones have appeared to be but slightly, if at all, related to ability in general, and to these the designation of special abilities or talents has been applied. Theoretically, artistic ability has been denoted as one of these, since "artistic aptitudes appear to be controlled primarily by special gifts rather than by general ability" (29, p. 505).[3] Actual evidence in support of this contention exists, however, only for that limited phase of artistic performance which has been designated as representative drawing. Furthermore, statistical data supplying such evidence of a low degree of relationship between general intelligence and ability in representative drawing have been derived, in the main, from investigations dealing with unselected groups, such as those of *Ayer* (2), *Elderton* (8), *Fischlovitz* (9), and *Ivanoff* (16). These studies have uniformly shown low, positive correlations between measures of the two variables.[4]

Complete agreement with these findings does not appear in the results of studies dealing with subjects selected for superior ability in art as manifested in other activities in addition to representation of single objects. The investigations of *Kerschensteiner* (17), *Kik* (18) and *Terman* (38) indicate that general intelligence is functioning in artistic performance and, to such an extent, that those showing marked accomplishment in art are invariably of high intelligence. A less extreme view in regard to the importance of general intelligence for successful performance in drawing, and one more compatible with the findings reported for unselected groups, is expressed by *Manuel*. He states that "general intelligence conditions the ability of drawers to acquire the advanced technique into which conceptual factors enter and to create original drawings of merit" (26, p. 119). It would seem logical that a certain degree of intelligence is necessary for marked achievement in any endeavor involving mental activity, but limiting the possibility of such achievement in art to individuals at the superior intellectual level does not appear tenable, if findings showing the character of special ability of representative drawing hold also for artistic performance in general. It is apparent from this summary of findings of the two types of investigations that studies using more comparable measures of artistic ability in unselected and in talented groups, with adequate samplings of each, must be made before any degree of consistency in results can be expected. For unselected groups no study has yet been made of the relationship between intelligence and the ability involved in creation of compositions judged for artistic merit, and similar studies with subjects at the high school and adult level who have demonstrated outstanding creative ability in art have been limited both in number and in scope.

---

[1] This study is part of the research program, Genetic Studies in Artistic Capacity, sponsored by the Carnegie Foundation for the Advancement of Teaching with financial aid from the Carnegie Corporation.

[2] The first section of this study (grade school subjects) was developed by Dr. Tiebout; sections two and three (special subjects and artists) chiefly by Dr. Meier.

[3] Quoted from Murphy's discussion of Spearman's theory of intelligence.

[4] Summaries of the findings of these studies have been made by *Goodenough* (12) and by *Hollingworth* (14).

The present investigation was therefore designed first, to determine the relationship existing between general intelligence, as measured by established intelligence tests, and artistic ability, as measured by the achievement of æsthetic quality in composition for a sampling of the child population in Grades I–VII inclusive; and second, to determine the status of general intelligence of two selected groups, one of artistically superior subjects at the high school level and the other of adult artists of recognized achievement in their profession.

II. Selection of subjects. A. Normal group. Subjects for the first phase of this study were 100 children in each of the grades from I–VII inclusive (106 in Grade II and 103 in Grade III), with an average IQ of 100 or slightly less, average chronological ages separated by about one year from grade to grade, and an approximately equal number of boys and girls, except in Grade I where the sampling gave twice as many boys as girls.

These subjects were drawn from the public schools of Des Moines, Iowa, a mid-western city of slightly over 140,000 population.[5] The majority of the subjects in Grades I–VI inclusive were enrolled in the Phillips School, but it was found necessary to supplement this group with pupils from the Hubbell School in Grades II–IV inclusive and the Perkins School in Grades V and VI. All of the Grade VII subjects were students in the Woodrow Wilson Junior High School, which serves the same area as the Phillips School. Complete information on the selection of these subjects is given in the preceding article (39) which is introductory to this and the studies of Saunders (33) and Meier and Stinson (28) in which these same subjects were used.

B. Artistically superior high school group. The artistically superior subjects used in the second phase of the investigation were selected by a canvass of the junior and senior high school population (Grades VII–XII inclusive) in Milwaukee, Wisconsin.[6] This locality was chosen because of the well organized department of art and the unusual facilities offered in securing subjects whose work had been carefully followed over a period of years. Also, the size of the population made it possible to exercise high selectivity and yet secure a sufficiently large sample for study. Attention had been called to the superior art work of a number of Milwaukee school children through reference to it in *The Art of the Child* (32). Certain of these children were identified as artistically superior and with this group as a nucleus, a list of subjects showing marked ability in art activities was made by the art supervisors. The subjects were divided into two groups, those designated as outstanding or unusual and those regarded as less outstanding but superior. Of the former group, 24 and, of the latter, 22 were included in this study. This second group was supplemented by 26 subjects rated by their respective art teachers as very good. Further study and rating of the art products of this group added by the teachers may later justify the inclusion of certain of these subjects in the first group, so that too rigid a classification was inadvisable.

---

[5] Acknowledgement is made to Miss Estelle Hayden, Director of Art in the Des Moines Public Schools, and to the principals and teachers co-operating in the study, particularly to Miss J. Ella Baker, principal of the Phillips School, where the greater part of the program was conducted.

[6] Acknowledgment is made to Mr. A. G. Pelikan, Director of Art in the Milwaukee Public Schools and to the art supervisors, whose co-operation made the program with the artistically superior subjects possible. Acknowledgment is also due to the superintendent, Mr. Milton C. Potter, for permission to test in the Milwaukee schools, and to the principals and art teachers at the various schools where the study was conducted.

These subjects represent a selection from nine high schools, two junior high schools and the 7th and 8th grades of six elementary schools. It had also been planned to study a group of artistically superior subjects selected in the elementary school grades in order that direct comparison of the findings could be made with results on the normal groups in these grades used in the first division of this study. However, since so few subjects had been chosen from any one of the 20 elementary schools, which were located over a large municipal area, time did not permit the testing of this group.

C. Adult artist group. For the study of the relationship between general intelligence and artistic aptitude in a highly selected group at the adult level, the co-operation of 200 contemporary American artists[7] was solicited through correspondence. The selection of these 200 artists was made by Dr. Meier on the basis of their outstanding achievement in art, as attested by the number and significance of honors won, and their general standing.[8] They were chosen from the approximately 5,500 names listed in the *Biography of American Artists* (1).

Of the total number solicited, 51 of the artists co-operated by taking the intelligence test submitted them with the request letter.

---

[7] This number does not include eight artists, letters to whom were returned, four because of incorrect address and four because the artist was deceased. Dr. Meier conducted the correspondence, including the original request letter to the artists and follow-up letters sent in certain instances.

[8] Because of his wide knowledge of artists and of the awards made in the art world, Dr. Meier was highly qualified to act as a judge of artistic achievement in making the selections for this study.

Of the remaining 149, replies were received from 31 artists expressing inability to do so for various reasons, chief among which were, lack of time, lack of interest or disapproval of assumed program, and illness. Similar reasons doubtless apply in the case of the majority of the 108 artists who sent no replies.

While it did not include all of the highest ranking artists among the number solicited,[9] this group of 51 artists was felt to be a fairly representative sampling of the large group of 200, not only in respect to rank in art but in respect to other important factors. The group was not limited to any one so-called school of art but represented a wide range of viewpoints. It also included artists of varying age, ranging from 28 to 82, with an average of 54 years, and in addition to painters, some of whom are also etchers, writers, or illustrators, there were among the number four sculptors and three caricaturists. Thus a wide range of interests within the larger field of art was represented.

III. Measurement of intelligence. For the grade and high school subjects the *Kuhlmann-Anderson* Intelligence Tests were used for the measurement of general intelligence. This choice was made because the *Kuhlmann-Anderson* series includes batteries of tests for the entire range of subjects which were to be studied, hence a uniform measure could be used throughout the grades. Also, the *Kuhlmann-Anderson's* use of the median mental age method of scoring in which separate mental age scores are assigned on each test in the battery "shows definitely and accurately on what particular tasks a child scores above or below his median mental age"(23, p. 4). Such information, it was thought, could be used in analyzing the relation of the general intelligence test results to the measures of artistic ability by utilizing the scores on the sub-tests.

The *Kuhlmann-Anderson* series was used in the study of the large samplings in Grades I–VII inclusive and in the study of the artistically superior subjects. In both instances it was administered as a group test in the classroom situation by Dr. Tiebout.[10]

---

[9] This point is discussed further in the description of rankings of these artists.

[10] In one of the Milwaukee schools, Peckham Junior High School, the *Kuhlmann-Anderson* test had already been administered hence the test was not repeated and IQs from the school record were used.

For the former group the *Goodenough* Drawing-of-a-Man Test was also given to a smaller number of subjects in Grades I–IV inclusive, for the purpose of making comparisons between the relationships between artistic status and the *Kuhlmann-Anderson* and *Goodenough* intelligence quotients. The *Goodenough* Drawing-of-a-Man Test was not given to Grades V–VII inclusive, nor to the artistically superior group since the test is more applicable to the lower age groups and norms are not given beyond age 13, which is below that of the majority of the superior group. For a number of the artistically superior subjects individual test results were secured on the *Stanford-Binet* Intelligence Test for the purpose of analyzing responses on the separate tests. Where two tests were administered, they were given on approximately the same date so that direct comparisons between IQs could be made, and consideration of possible changes in IQ over a period of time would not be necessary.

Adult artist group. In the case of the adult artists the *Otis* Self-Administering Tests of Mental Ability, Higher Examination, Form A, were selected for the measurement of intelligence. This series of tests was chosen because, as the name implies, it is 'self-administering,' the function of an examiner being merely to see that the subjects understand the printed directions given on the test form and to give the 'go' and 'stop' signals. Since the investigation with the artists was conducted entirely through correspondence it was impossible to have an examiner present to control the time limit and the conditions under which the test responses were made. This had to be left to the individual artist. Definite information was given in the correspondence in regard to control of both these factors and the artist's attention was particularly called to the absolute necessity for rigid adherence to instructions given on the test blank, the time limit being heavily underlined in red for especial emphasis.

Not only was it felt that this test could be completely self-administered with considerable success by adults, but it was thought that the short 30 min. time limit would be more conducive to co-operation on the part of the artists than one making greater demands on their time. Furthermore, this test represents a most significant prognostic and vital short collection of items, doing virtually the work of longer tests.

IV. Measurement of artistic ability. For the measurement of the second factor, artistic ability, no standardized tests were available as in the case of general intelligence. The term artistic ability as here used refers to ability to produce or, as defined by *Kinter* (19, p. 8), " the ability to create forms of artistic merit," in contradistinction to artistic appreciation. With the large number of subjects, certain limitations were of necessity placed upon the extent to which art production could be sampled. Since the aim was to measure a type of ability other than that dealing exclusively with accuracy in representation, which had been studied by previous investigators, it was decided to have each subject make three paintings, or compositions, illustrating stories, to be judged on the basis of the æsthetic quality achieved. Since the only available scale for such measurement, the *Kline-Carey* Measuring Scale for Freehand Drawing, Part II (21), did not include sufficient samples to give the desired differentiation in products, particularly at the lower grade levels, a new measuring series was devised consisting of separate scales for each of the seven grades. The development and use of these scales in determining the artistic status, or painting score, of the subjects used is described in the preceding article (39). In brief, the final painting score of each subject represents the average value of his three paintings as judged by the scale devised for his grade.

Ranking of adult artists. The procedure in determining the artistic standing of the selected high school subjects and of the adult artists has already been discussed. In order to interpret results in the case of the adult artists, a further step was taken to determine more exactly their ranking in their profession. The names[11] of 49 of the artists co-operating in the study by taking the *Otis* Test were submitted to a number of judges, who were requested to rank the artists in the order of their importance from the foremost artist in the list to the least important.[12] In determining the relative importance of these artists the judges were instructed to consider the significant honors won and the probable permanent value of the work of each. The judges were also advised to emphasize the artist's standing as a "creative artist rather than as a merely skillful technician or a clever painter of a given type of subject-matter such as portrait or landscape" (quoting from correspondence with the judges by Dr. Meier).

---

[11] A total of 57 names appeared on the list, since eight artists who might respond to a second request were added. Two of the 51 artists returned their test responses after the list had been sent to the judges.

Judges were chosen who were qualified to make such rankings because of their knowledge of the more important contemporary American artists in their capacities either as art museum directors, directors of art education, or art editors. Of the four judges who made rankings of the artists, one submitted a list omitting only three of the 49, while a second omitted 18. The rankings of the other two judges were too incomplete to serve the purposes of this study. While more or less complete rankings were thus obtained from only two judges, their unusual qualifications would warrant the attachment of a great deal of weight to their choices. The correlation between the rankings of the two judges was .56 as computed by the method of rank-differences (**10**, p. 190). This is equivalent to a product-moment r of .58±.08. Both rankings were used for correlation with other data in the treatment of results.

In addition to ranking the artists whose names appeared on the list, both of the judges submitted supplementary lists of important artists who had not been included. One listed 17 artists whose respective ranks would fall between 4 and 27 in comparison with those on the original list. Another named 13 artists whom he felt should be included among the most important in the country. It is therefore apparent that results were not secured from all of the highest ranking artists, as has already been pointed out, but the fact that these supplementary lists are not more extensive would suggest that the upper extreme is more or less adequately represented.

---

[12] The selection of the judges and all correspondence with the judges were taken care of by Dr. Meier.

V. Data on Subjects. A. Normal group. Results obtained through the use of these measures of intelligence and artistic ability described in the preceding section, together with other necessary data, were compiled in tabular form for all groups of subjects used in this study. Table I presents summarized data for the entire group of subjects in Grades I–VII including, in addition to the material reproduced from Table I of the preceding article (39), a column giving the average of the painting scores as the measures of artistic ability have been designated.

Table I. *Summarized data on normal group*

		Sex		CA		IQ (K–A)			
Grade	No.	M	F	Av.	σ	Av.	σ	Av.	σ
I	100	67	33	6–8	5	100.38	10.53	2.70	.57
II	106	46	60	7–9	6	98.96	8.10	2.74	.65
III	103	51	52	8–9	7	98.78	9.99	2.64	.56
IV	100	49	51	9–10	9	99.82	11.28	2.66	.56
V	100	48	52	10–9	9	99.03	10.59	2.35	.57
VI	100	57	43	11–10	8	98.16	10.56	2.44	.64
VII	100	54	46	12–8	8	99.39	9.90	2.04	.51

Since the results of the subjects in Grades II–VI inclusive of the Phillips School have also been treated apart from those for the entire group, data for this group alone are given in the following table. Grades I and VII are not included since these groups, as has been pointed out, were composed entirely of students in the Phillips and in the Woodrow Wilson Schools, respectively.

Table II. *Summarized data on normal group, Phillips School*

		Sex		CA		IQ (K–A)			
Grade	No.	M	F	Av.	σ	Av.	σ	Av.	σ
II	78	33	45	7–8	6	97.68	7.80	2.63	.80
III	75	38	37	8–7	6	96.57	9.12	2.45	.46
IV	70	33	37	9–10	5	98.31	11.49	2.56	.54
V	70	31	39	10–10	8	97.02	11.16	2.13	.51
IV	73	40	33	11–10	8	96.06	10.05	2.25	.58

Table III gives similar data for those subjects in Grades I–IV inclusive of the Phillips School who took the *Goodenough* Drawing-of-a-Man Test in addition to the *Kuhlmann-Anderson* series. Averages and sigmas for both IQs are given, 'K-A' referring to *Kuhlman-Anderson* and 'G.' to *Goodenough.*

TABLE III. *Summarized data on subjects taking Goodenough Drawing-of-a-Man Test*

		Sex		CA		IQ (K-A)		IQ (G)			
Grade	No.	M	F	Av.	σ	Av.	σ	Av.	σ	Av.	σ
I	84	58	26	7–0	2	101.35	11.20	104.40	10.95	2.74	.58
II	65	28	37	8–2	6	97.86	7.74	99.82	19.32	2.70	.67
III	70	37	33	8–7	6	96.60	9.30	105.35	14.10	2.48	.42
IV	58	29	29	9–9	7	98.35	11.95	103.30	16.85	2.54	.56

B and *C*. *Artistically superior high school subjects and adult artists.* For the artistically superior or 'X' subjects selected from Grades VII–XII inclusive in Milwaukee, individual records were compiled.[13]

VI. Results of study. *A. Normal group.* To determine the relationship existing between results of the measures of general intelligence and of artistic ability for the normal group of subjects, correlation techniques were employed. Results of this statistical treatment of the materials are presented in the following series of tables. Table IV gives the *Pearsonian* coefficients of correlation between the *Kuhlmann-Anderson* IQ and painting score and, in the last column, the partial correlations between IQ and painting score with CA held constant. Correlations between painting score and chronological age and between CA and IQ are also given to show the derivations of the partial r's. It was necessary to rule out the factor of CA for a more accurate representation of the relationship between the two measures under consideration, since the subjects were grouped by grades, rather than by chronological age.

Table V gives the same data for the results of Phillips School taken separately and Table VI for the results using the *Goodenough* IQ. Results from Grades I and VII are not repeated in Table V since they appear in the preceding table.

An examination of both the correlations between painting score and IQ and the partial r's given in Table IV reveals a range

---

[13] These are on file as records of the Division of the Psychology of Art. The artists' records were obtained under conditions of a confidential nature and are consequently not available.

TABLE IV. *Kuhlmann-Anderson IQ and painting score*

Grade	Ptg. S & IQ	Ptg. S & CA	CA & IQ	Ptg. S & IQ (CA constant)
I	.16±.07	.24±.06	—.52±.05	.35
II	.21±.06	.18±.06	—.47±.05	.34
III	.49±.05	.12±.07	—.21±.06	.53
IV	—.05±.07	—.02±.07	—.46±.05	—.07
V	.17±.07	.13±.07	—.51±.05	.28
VI	.22±.06	.01±.07	—.63±.04	.29
VII	.01±.07	.10±.07	—.64±.04	.02

TABLE V. *Kuhlmann-Anderson IQ and painting score, Phillips school*

Grade	Ptg. S & IQ	Ptg. S & CA	CA & IQ	Ptg. S & IQ (CA constant)
II	.23±.07	.22±.07	—.61±.05	.47
III	.29±.07	—.06±.08	—.46±.06	.30
IV	—.07±.08	—.11±.08	—.57±.05	—.16
V	.01±.08	.25±.08	—.58±.05	.19
VI	.01±.08	.01±.08	—.59±.05	.09

TABLE VI. *Goodenough IQ and painting score*

Grade	Ptg. S & IQ (G)	Ptg. S & CA	CA & IQ (G)	Ptg. S & IQ (G) (CA constant)
I	.31±.07	.30±.07	—.22±.07	.40
II	.12±.08	.23±.08	—.42±.07	.24
III	.41±.07	.00±.08	—.12±.08	.39
IV	.18±.09	—.03±.09	—.30±.08	.18

from a slightly minus correlation of .07 in Grade IV to a positive correlation of .53 in Grade III. These figures fall approximately within the range of slightly above 0 to .50 correlation reported by previous investigators dealing with ability in representative drawing in unselected groups. Moreover, they approach very closely the range in correlations of .140±.070 to .510±.056[14] found by *Bird* (3) whose recent study corresponds most nearly to the present one in that a standardized intelligence test was employed and an approximately equal number of subjects in a comparable, though slightly larger, age-range was used. Practically all of the earlier studies were limited in their measures of intelligence to teachers' estimates of general ability or to standing in school subjects, since intelligence tests had not yet been developed. Also the age-range studied was narrow in most cases and in some instances the number of subjects in any one age-group was too small for reliable results. *Bird's* study differed from the present one mainly in the measure of artistic performance used, which resembled that of the earlier investigations. *Bird's* subjects made three drawings of a cat running after a ball, the first from memory, the second from a model, and the third from memory after the model had been removed, these drawings being scored for accuracy of representation and the portrayal of action.

---

[14] The correlations cited are those for combined score on three drawings and Dearborn IQ.

While *Bird* interprets the correlations found in his study as showing that intelligence is one of the factors which influences ability to draw, he points out that the degree of relationship between the two factors is slight when intelligence is measured by the *Dearborn* Group Test.

In the light of the agreement between the correlations in the present study and those of previous investigators, particularly of *Bird*, apparently no greater relationship to general intelligence exists for the type of artistic ability considered in this study than for ability in representative drawing, the relationship in either case being generally positive but low.

Further examination of Table IV shows that with the exception of Grade IV a low positive correlation between CA and painting score combined with a more marked negative correlation between CA and IQ effected an increase in correlation between painting score and IQ, when the influence of age was eliminated by partial correlation. However, since these changes in correlation were comparatively negligible in Grades III, IV and VII and not considerably larger in the remaining grades, such age-differences as existed within a grade seem to have had little effect. The most marked increase occurred in Grade I, for which the correlation between CA and painting score was higher than in any other grade and which was also one of the two grades where this correlation was higher than that between painting score and IQ. This is to be expected, since the younger children in Grade IB have had much less experience with the art media used than the older children in Grade IA; many of them have made their first paintings on their entrance to school.[15] The same general tendency to increase somewhat manifests itself in the results of partial correlations for Phillips School taken separately (see Table V). Almost the same degree of increase is noted for Grade II as in the case of Grade I, probably due to the operation of similar factors.

In addition to ruling out the factor of CA, some check on the factor of art training was necessary and for this purpose correlations for the Phillips School are presented separately. In spite of the fact that the general course of art instruction is uniform throughout the Des Moines Public School system, possible variation in the training in different schools may have been operating to affect the correlations in Grades II–VI inclusive where the subjects at Phillips School were supplemented by a smaller number of subjects from a second school. Using Phillips School only, the training factor, although not rigidly controlled, is more constant for all subjects within each grade.[16]

A comparison of Tables IV and V shows that in general the correlations for Phillips School agree with those for the entire group in being low and positive in all grades except IV, where the results are again slightly negative. The correlations are, however, somewhat lower in each instance, with the exception of Grade II where the product-moment correlation is higher by two points. Partial correlation raised the correlation in this grade considerably more in the Phillips School than in the entire group, so that the difference in the partial r's is larger. The lowering of the correlations in Grades III–VI inclusive may have resulted in part from the reduction in the number of subjects and the limiting of the IQ range at the upper level with the use of Phillips subjects only, but an examination of the actual paintings indicates that certain aspects of the training in the schools from which the higher IQ children were drawn were of such a nature as to cause these children to make higher painting scores. The Hubbell subjects in Grades II and IV, particularly in the former, had apparently been trained more specifically in space-filling and hence made more complete and unified compositions. In the case of the Perkins School, Grade VI and to some extent Grade V, the paintings possessed certain unusual color qualities, the achievement of which doubtless resulted from special training in this respect. Only in Grade II, where such differences in training effects are not observable in the products of the two schools, was the product-moment correlation between painting score and general intelligence not lowered and only slightly altered. These Phillips correlations, together with the almost zero correlation in Grade VII, suggest that with a group of subjects comparable in number and IQ range to those from the combined schools, but with the factor of training more adequately controlled, lower correlations between painting score and IQ than those given in Table VI would result in the grades above II.

---

[15] The fact that two crayon drawings in addition to two paintings were made by the subjects in Grade IB did not operate to lower their scores, since the average values assigned to two paintings approximates that of the average for both the crayons and paintings which was used as the score of each subject.

[16] Regular art classes taught by an art teacher are introduced at Grade IV. In the earlier grades the art activities are related to the teaching program and are directed by the regular classroom teacher.

Further examination of the partial correlations secured with the Phillips School data reveals that, while all of the correlations are low, comparatively higher correlations are found in the first three grades than in the upper grade, with a decided drop evident at Grade IV. *Goodenough* has pointed out that there is a close relationship between general intelligence and concept development as shown in drawing in children up to the ages of about ten, after which there is a decrease in the relationship, drawing taking on the character of a specialized ability rather than a general means of expression. While a high relationship to general intelligence is not found in any instance in this study which deals with a type of ability distinct from that to which *Goodenough* refers at the younger ages, the grade range in which the correlations are of the larger magnitude (I–III) parallels the ages at which *Goodenough* finds a high relationship to exist. Thus while a close relationship does not exist between ability to achieve æsthetic quality in compositions and general intelligence, as is the case with drawing used as a general means of expression at the younger ages, there is a tendency toward some relationship in the lower grades in contrast to the upper grades where the correlations are approximately zero. Apparently even the type of ability measured in this study does not emerge as a special ability, almost entirely independent of general intelligence, until around the ages of nine and ten.

When the measures of artistic ability used in this study were correlated with IQ's obtained from the Drawing-of-a-Man Test devised by *Goodenough* to measure graphic concept development, low, positive correlations were again found in Grades I–III inclusive, and also a slightly positive correlation in Grade IV, instead of the slightly negative correlation secured with the *Kuhlmann-Anderson* IQ in this grade. Since these correlations on the whole do not vary in any considerable degree from the correlations between the *Kuhlmann-Anderson* IQ and painting score, it appears that the *Goodenough* Test, a performance test measuring the developmental characteristics in children's drawings of a man, does not relate any more highly to the child's ability to achieve æsthetic quality in composition than does the usual verbal type of intelligence test such as the *Kuhlmann-Anderson* series.

no1psychologicalm48ameruoft

30005030332798

Bird (3) reports a correlation of .49±.03 between combined drawing score and IQ on the *Goodenough* Drawing-of-a-Man Test for a group of 248 subjects, the majority of whom were between 6–6 and 9–6 years, an age-range comparable to that of the subjects in this study. The higher correlation obtained by *Bird* is probably due to the fact that the drawing ability, which he measured, is more closely related to the aspects of the drawings of a man measured by *Goodenough* than is the measure of æsthetic quality in paintings used in the present study. Still higher correlations ranging from .63 to .86 were obtained by *Manuel and Hughes* (27) who correlated IQ's on the *Goodenough* with ratings of the test responses for their quality as drawings, the records of 836 children in Grades I–IV inclusive being used. Furthermore, the correlations in the first and second grades were higher than in the third and fourth grades. The authors suggest that probably the factors which differentiate talent in drawing are less obvious in the first grade than they are in later grades. While the *Thorndike* Scale used as the basis for ratings of the drawings is limited in its adequacy for this purpose, it is probably true that there is little to differentiate performance as measured by *Goodenough* and the artistic quality shown in a single drawing of a man, particularly in the earlier grades. Use of the phrase 'talent in drawing,' however, in this instance is rather misleading because of the decided limitation in the measure of this factor.

On the whole in the earlier grades, drawings of a man offer very little which is measurable as artistic quality apart from certain aspects of the drawings taken into account by the *Goodenough* Test. The *Kline-Carey* Scale for a man running (20), for example, aside from the factor of the portrayal of action, lays stress on the inclusion of the different parts of the body, the correct attachment of arms and legs, the proportions achieved, *etc.*, although, of course, not going into such detail nor dealing with developmental sequences as *Goodenough* does. Additional features, such as light and shade, do not appear until the later samples of the scale, representing work in the upper grades. *McCarty's* Scale (25) for drawings of human beings, devised for ages 4–0 to 8–6 yrs., is still more closely related to the *Goodenough* Drawing-of-a-Man Test in the type of ability measured, dealing as it does with ability to express ideas vividly, clearly and accurately rather than with artistic merit. The low relationship between the *Goodenough* IQ and painting score in this study should be expected, however, since emphasis has been placed on the achievement of certain æsthetic qualities in paintings, rather than on realistic representation, where certain aspects of measurement in common with the Goodenough may be present to increase the degree of correlation.

Correlations between the *Goodenough* and *Kuhlmann-Anderson* IQ's were as follows: Grade I, .43±.06; Grade II, .42±.07; Grade III, .42±.07, and Grade IV, .19±.07. In view of the .74±.02 correlation reported by *Goodenough* between IQ on the Drawing-of-a-Man Test and *Stanford-Binet* IQ, higher correlations would be expected and an explanation of the discrepancy is not apparent. These correlations for Grades I–III inclusive agree closely, however, with the r of .47±.03 reported by

Bird (3) for the *Dearborn* Group Tests of Intelligence. They also agree with *Goodenough's* correlations with teachers' judgments of ability, which were found to be .44 within the first three grades, with a drop to lower correlations thereafter.

VII. Results of study. B. High school subjects. For the group of artistically superior high school subjects used in the second phase of the study, intelligence test results give an average IQ on the *Kuhlmann-Anderson* series of 109.35 in the case of the 24 starred or unusual subjects and of 107.30 for the remaining 48 subjects. These data show that the artistically superior also tend to be somewhat superior in intelligence, the group average falling at the upper limit of average IQ's, according to the classifications given by *Terman* (36). Rank in intelligence apparently gives little indication of the degree of artistic superiority, however, since the 24 starred subjects have only a slightly higher average IQ than the remainder of the group. Within the group of starred subjects, also, the subject indicated as the most outstanding had an IQ of only 101 on the *Kuhlmann-Anderson* series and of 105 on the *Stanford* series. On the basis of their IQ's one of the group of 24 outstanding subjects would be included in the *Terman* classification of dull; ten, average; eight, superior, and five, very superior. While there is a slight preponderance of these subjects in the superior and very superior groups it is obvious that an average IQ does not preclude the possibility of successful artistic performance at the high school level.

These findings are at some variance with the results reported by *Terman* in his investigations of gifted children. *Terman* attempted to locate children of only average or moderately superior IQ who were outstanding in certain special abilities, including art, but in a school population of a quarter-million only 26 children (15 in art) were found who showed any unusual promise. A 'follow-up' of these subjects showed that the early promise was not fulfilled in a single case. *Terman*, therefore, concludes that "without superior general intelligence, special ability in music and art inevitably falls short of really great achievement. All of the young musicians and artists of genuine accomplishment whom the writer has studied," he adds, "have had without exception, high intelligence quotients."

In considering *Terman's* findings it should be pointed out that the degree of artistic ability of the children originally selected for the special ability group is somewhat in question. Although great effort was made to locate these special cases, with the final selection made from a special art class of supposedly talented children, satisfactory measuring methods were not available and *Terman* states that "with two or three possible exceptions there was little evidence of anything more than 'copying ability'" (37, p. 34). On the face of this evidence there is room to doubt that the special ability group selected would be expected to achieve any marked success. The fact that a larger number of children with creative ability and with an IQ below the very superior level could not be located is difficult to explain except perhaps on the ground that adequate means for the selection of such subjects were not available. It is also possible, as *Terman* points out, that such subjects could have been more easily identified at a somewhat higher age-level than that which was used. The finding that certain of the intellectually gifted were markedly successful in art, among other things, would be expected, since there are without doubt those children in whom high intelligence, combined with artistic ability, contributes toward outstanding achievement in this specialized activity. As has been pointed out, instances of artistic superiority combined with high intelligence, although not as high as that possessed by *Terman's* intellectually gifted, are also found in this study. The present study also shows, however, that the artistically superior need not necessarily be intellectually superior.

Kerschensteiner and *Kik*, who report findings similar to those of *Terman*, seem to have located children demonstrating definite creative, rather than copyist, ability in art. Their conclusion, that great talent for graphical expression is regularly connected with good intellectual endowment, cannot be adequately interpreted, however, for purposes of comparison with the results of studies using intelligence tests, since intellectual endowment was indicated to them solely by the child's status in general school work.

While results for the artistically superior subjects in this investigation differ from those of *Terman,* and perhaps of *Kerschensteiner* and *Kik,* this disagreement is probably due to the limitations pointed out in these earlier studies. In the present investigation findings are based on measurements with a standardized intelligence test of a large number of subjects selected for their artistic superiority in creative, rather than copyist, activities. It is fairly improbable that the average IQ found was unduly weighted by a selection of artistically superior subjects from schools of predominantly lower or of predominantly higher than average IQ level, since practically all junior and senior high schools in the Milwaukee system were used. Hence, the population from which the subjects were drawn would be expected to show a normal distribution of intelligence. Furthermore, the fact that the subjects were in the junior and senior high school grades would give increased surety of their artistic superiority, since there had been a longer period for observation of their work, than in the case of younger pupils, and greater opportunity for proof of their consistency in artistic performance.

The results of this phase of the study dealing with the artistically superior indicate the presence of a somewhat higher relationship with general intelligence for the upper extreme of the distribution of artistic ability than for the normal group, if correlations similar to those found for the latter in the upper elementary school grades hold also for the high school level at which the talented subjects were selected. While no data are available in the study to explain this difference, it is likely that a certain degree of general intelligence is required to profit from the training necessary for the type of achievement demonstrated by the artistically superior at these grade-levels, with the result that average and above average IQ's are found in the majority of cases in the selected group. The results of *Manuel*'s study of 19 artistically talented subjects, ranging from fifth grade to university students, furnish some basis for this inference in regard to the relation between intelligence and development in drawing, although his evidence bearing directly on this point seems to be limited to the finding that one subject of inferior intelligence seemed unable to profit rapidly by advanced instruction in technique and that another subject of markedly inferior intelligence showed less originality than ability in graphic representation.

VIII. Results of study. C. Adult artists. With the extension of the second phase of the study to the highly selected group represented by the adult artists, results again show a higher than average degree of intelligence to be co-existent with artistic superiority, the superiority in this instance representing the extreme of genuine accomplishment among those following art careers. The average IQ of the group of 51 artists as measured by the *Otis* Self-Administering Tests of Mental Ability was 118.43, which approaches the 120 IQ set by *Terman* (36) as marking the division between superior and very superior intelligence. Since this average is nine points higher than that for the 24 junior and senior high school subjects showing outstanding ability in art, indications are that a still higher level of general intelligence is necessary for eventual marked achievement in art. As with the younger group of selected subjects, however, a wide range in IQ within the group is found, there being nine artists of average intelligence, 12 of superior, 26 of very superior, and one of 'near' genius or genius intelligence (according to *Terman's* classification). With the adult artists further study of the nature of the relationship between artistic achievement and status in general intelligence was possible through experts' ranking of the artists in the order of their importance. Correlations between rank in art assigned by Judge 1 and rank in intelligence, using the rank-difference method (10, p. 190) was —.12 and for rank in achievement by Judge 2, .10, the equivalent product-moment r's being —0.13±.10 and .12±.12, respectively. These correlations indicate that within the group there is no relationship between the individual's rank as an artist[17] and his rank in intelligence. Thus it would seem that while a high degree of intelligence characterizes the successful artist group as a whole, intelligence level within the group may give little or no indication of the individual's importance as an artist. Since many factors related to general intelligence doubtless operate in the attainment of eminence in art, as in other fields, it is not surprising to find a superior IQ level for the group. The 'special ability' character of the art talent factor, however, is supported by the absence of relationship between rank in importance as an artist and IQ.

---

[17] A striking observation in this connection is that about ten names appeared consistently in the upper ranks of all four lists.

The possible effect of differences in chronological age on these results was checked through correlation with rankings of the artists and with IQ. The rank-difference correlation between chronological age and rankings by Judge 1 was —.37 and for Judge 2 was —.17, the equivalent r's being —.39±.09 and —.18±.13, respectively. These correlations show a tendency for the judges to give higher rankings to the younger artists, indicating perhaps that of the more important artists greater co-operation was secured in this study from the younger than the older group. A —.14±.09 correlation between chronological age and IQ also indicated a tendency, though slight, for the younger artists of the group to score higher in intelligence. The simultaneous operation of the tendencies for the younger artists of the group to score higher in intelligence and to be ranked higher in achievement would have tended to raise the correlation found between rank as an artist and IQ. This would not alter the interpretations of the relationship between rank as an artist and status in general intelligence already given, however, since indications are that the low correlations of —.13 and .12 are even somewhat higher than would have resulted if the factor of chronological age had been eliminated.

The influence of certain other factors on the results cannot be checked because of the difficulties inherent in the manner of conducting investigations of this type in which direct testing of the entire group chosen for study is impossible. The extent to which the IQ distribution, found with the group of 51 artists, represents that of the larger group solicited cannot be known since some degree of selection relative to the factor of intelligence may have been operating to determine the group participating. Furthermore, since the intelligence test was in each instance entirely self-administered it is probable that the conditions under which the test was taken were not equally favorable or well controlled for those who responded, as would be more nearly the case if a direct administration of the test had been possible.

Possibilities for a more complete interpretation of the intelligence test results of the adult artists are also limited by the fact that similar data from groups of outstanding men in other professions were not secured for purposes of comparison. *Hollingworth* (14, p. 152) has expressed the opinion that "probably those who achieve eminence in the arts are, on the whole, as highly endowed with general intelligence as are those who win eminence in other kinds of careers." On the basis of the results for the artists in this study, however, it appears that the average of intelligence found in this group is lower than that which would be expected for similar groups in certain other professions, at least in those in which achievement is more obviously dependent on the types of mental activity measured by intelligence tests. Little direct evidence in connection with this problem is available in the literature. *Cox* in a study of the early mental traits of 300 geniuses found the rating in intellectual traits for a group of 13 artists to be significantly lower than that for eminent men in general. Since this study was made in retrospect, however, these findings are based on assumed IQ's and are hardly comparable to results in the present study. While the classification, artist, does not appear in the familiar occupational intelligence ratings from the Army Alpha Test results (30, p. 829), one group of 28 artists appears in an earlier table in such a rank as would indicate an intelligence falling at the lower limit of the 'B' or superior group, below the occupations identified as engineer officer, medical officer, civil engineer, and a number of other professions. Since this investigation was quite limited in many respects and dealt with comparative status among occupational groups without regard for rank within the group, it gives only an indication by inference that artists of marked achievement would be somewhat lower than men of similar achievement in the careers mentioned.

For the field of music, experimental data on the relationship between intelligence and achievement have been contributed by the extensive research conducted at the Eastman School of Music (35). While the problem was approached from a different angle than that of the present study, these results in music offer possible inferences for the situation in art, since both are regarded as special abilities. After considerable experimentation with prognosis of musical achievement at the Eastman School, it was found best to combine the results of the Seashore Measures of Musical Talent with the Iowa Comprehension Test (a college 'intelligence' test). In the cumulative key developed for this purpose, a series of score combinations are designated for each of five categories of prediction, ranging from 'discouraged' to 'safe.' These show for the prediction of 'safe,' intelligence ratings of superior, very good and high average combined with a superior talent profile. With the artist group in this study IQ's within approximately the same range were found to be present in the case of presumably superior artistic talent, although there were no actual measures of this factor. The fact that intelligence conditioned the use of musical potentialities to such an extent that a measure of this factor was included for prediction of success in the music school courses lends confirmation to the explanation advanced for the superior IQ average of the artists and the high average IQ of the selected high school group, since intelligence probably functions in a similar manner in achievement in art.

Analysis of intelligence test responses. In an attempt to discover whether certain types of tests included in the different intelligence tests were more highly related to artistic ability than others, analyses were made of each, treating scores on the component tests or items separately. In the case of the *Kuhlmann-Anderson* series, as has been pointed out, comparisons may be made between standing on the different tests in the battery since a mental age is assigned to each. The number of above and below median mental age responses made on each test by a group of subjects, whose painting scores were above one sigma of the distribution in Grades III and IV of the Des Moines group,[18] was therefore recorded in order to determine whether subjects of higher rank than the average in artistic production tend to score above or below their median mental age more frequently on certain tests than on others. On examination of the mental ages on the separate tests taken by the subjects in each of these grades, however, it became apparent that there was a general tendency for the group as a whole to score higher on certain tests and lower on others. Similar records for a group of subjects whose paintings were below one sigma of the distribution and for the larger group of subjects falling between ±1 sigma, were therefore compiled. With these data for comparison results for the artistically superior group did not vary to any considerable degree from those for the other two groups, the differences in any case being slight and probably insignificant. With the type of tests included in the *Kuhlmann-Anderson* series it appears that relationship to artistic ability is neither more nor less marked in the performance on one test of the series than on another.

For the group of 23 artistically superior subjects in the Milwaukee Schools, to whom *Stanford-Binet* Intelligence Tests were given, results were recorded by separate tests for purposes of analysis. Of the responses of the 22 subjects who had a basal age below year 16, the number of 'passes' on the individual tests of this age series was as follows: vocabulary, 6; interpretation of fables, 14; differences between abstract terms, 10; problem of enclosed boxes, 12; repeating six digits reversed, 12, and using a code, 16. For the 20 subjects passing one or more tests at year 18, there was the following number of passes: vocabulary, 3; *Binet's* paper-cutting, 14; repeating eight digits, 8; repeating the thought of a passage, 13; repeating seven digits reversed, 8, and ingenuity, 7. The largest number of 'passes' in year 16, as will be noted, was on the 'using a code' test and in year 18, on *Binet's* paper-cutting test. For the latter test, *Terman* states that "success seems to depend upon constructive visual imagination" (36, p. 339). In the case of the former, while he points out that "the significance of the test must be expressed in other terms than the kind of imagery it may happen to bring into play," he states also that "nearly all subjects find it necessary to trace the code form in imagination, from the beginning up to each letter whose code form is sought" (36, p. 331). Thus these two tests, particularly the *Binet* paper-cutting test, appear to tap certain functions which would be expected to be more important in artistic production. It should be pointed out, however, that there were almost as many 'passes' on interpretation of fables in year 16 and repeating the thought of the passage in year 18, in which the mental abilities most involved would seem to have little direct implication for functioning in art. In the light of the oft-expressed opinion that the artistic individual is visually rather than verbally minded, it is interesting to note that the vocabulary test was more often failed at both levels than any other test.

---

[18] A similar procedure was not followed in Grades I and II because of the presence of a number of maximum and minimum scores and was not carried beyond Grades III and IV because results did not seem to warrant its continuance.

This analysis of test responses is of course limited by the small number of subjects under consideration and by the fact that similar analysis of the responses of an unselected group have not been made. *Skalet* (34) has made such a study but gives only a tentative interpretation of results above the nine year group because of the limited number of cases used above this age level. Also, it is difficult to analyze the mental functions operating in performance on any particular test, and no extended measure of any function is given in a single sub-test as such was not intended in the development of the general intelligence test. Tests of special aptitudes, such as those recently developed by *Thurstone*, would have been more valuable for analytical purposes, but these are devised for the college level and are also quite time-consuming in their administration.

For the adult artists, responses to each of the 75 items of the *Otis* Test were recorded separately and the number of failures on each item noted. Although *Otis* states in the manual of test instructions (31) that the items are arranged in order of difficulty, no data are given as to the exact difference in difficulty between items. No weighting of the items to equate this factor is therefore possible for purposes of comparison. The grouping together of certain items of varying degrees of difficulty but dealing with similar materials, however, shows that a greater number of errors occurred in those items involving following directions and number concepts, particularly the latter. Items which have been classed as following directions deal mainly with locating the correct letter in a series by taking certain steps given in the directions, and those identified as dealing with number concepts involve the finding of the wrong number in a series of numbers. The greatest number of errors for a single item occurred on number 68 which deals with *line and space* concepts. This may be explained, however, by the fact that this problem is more of a geometrical (mathematical and abstract) than one of visual space relations. This item, as well as items 66 and 67 on which there was a large proportion of errors, would be included with number concepts in the larger classification of mathematical abilities. While it thus appears that a majority of the items presenting difficulty for the artists involve mathematical abilities, items in a third group in the same classification, arithmetical reasoning, are missed no more often than others of approximately the same difficulty.

Of the more difficult items, the smallest number of errors was made on items 71 and 69, which fall in the large classifications of verbal and of reasoning abilities, respectively. No tendency to fall into more or less specific groups is apparent, however, in the case of the items showing the greater number of correct responses, as was true for the items most frequently missed. It appears that the test does not include items particularly diagnostic of the types of abilities more peculiar to the artist's equipment.

IX. Summary and conclusions. This study was devised to determine the relationship of general intelligence to artistic ability, both in the degree to which it is present in the child population and as it is manifest in superior performance at the high school level and in genuine achievement in the artist. The entire study was planned in the light of previous findings which had indicated the following considerations to be of major importance, namely, age of the subjects, type of artistic ability measured, and nature of the group studied, whether unselected or composed of the artistically talented. The problem was therefore approached with the view of determining the effect of these various factors, by considering both selected and unselected groups within a wide age-range, measured for a similar type of artistic ability. Data on representative drawing collected by previous investigators for unselected groups served for purposes of comparison in regard to the latter factor.

For the first phase of the study each of the school grades, from I to VII inclusive, was represented by a sampling of 100 subjects, normally distributed as to intelligence, with an IQ average approximating 100. The artistic ability of these subjects was determined by their average scores on three paintings, judged for æsthetic quality by means of scales devised for this purpose.

For the first of the selected groups included in the second phase of the study, a canvass of the junior and senior high school population in a large mid-western city resulted in the rating of 72 students as artistically superior by the art supervisors and teachers, the former designating 24 of this number as outstanding. The second group, which represents a high degree of selection at the adult level, was composed of 51 artists of recognized achievement in their profession. Two expert judges ranked the artists in order of importance.

Measurement of the factor of general intelligence was based in each instance on responses to a standardized intelligence test, the *Kuhlmann-Anderson* in the case of the normal and selected high school groups and the *Otis* Self-Administering Tests of Mental Ability in the artist group. The *Goodenough* Drawing-of-a-Man Test was also used for certain subjects in the first group.

A. *For the normal group (grade school subjects).* On the basis of correlations between the measures of intelligence and of artistic ability, the following conclusions are drawn:

(1) Only a slight relationship is found between general intelligence and artistic ability in the grades above III, where

the correlations are low and unreliable in each instance, being slightly negative in Grade IV. For the lower grades the relationship is significant for Grade III, and partial correlations, holding CA constant, show an increase in relationship in Grades I and II, indicating a positive but only a fair relationship between general intelligence and artistic ability at these years (ages 6 to 8).

(2) The relation of general intelligence to artistic ability, functioning here in *free or creative expression*, appears to be virtually the same as that found by other investigators for ability in *representative* drawing.

(3) A *high* relationship to general intelligence is not found in Grades I to III inclusive which parallel the ages in which development in drawing, regarded as a general means for expression of emerging concepts, has been shown to be highly related to general intelligence as contrasted to the latter ages in which the importance of this factor decreased with a change in the nature of drawing. The present results agree, however, to the extent that a low to fair degree of positive relationship is manifest in these earlier grades.

(4) For the *Goodenough* Drawing-of-a-Man Test a low degree of relationship to artistic ability exists as indicated by low, positive correlations in Grades I to IV inclusive. Since this finding is in general agreement with that for the *Kuhlmann-Anderson* test in these grades, particularly I to III inclusive, it appears that for the type of artistic ability considered in this study there is no greater relationship to an intelligence test based on graphic conceptualizing (the *Goodenough*) than to the usual verbal type of intelligence tests. The fact that a higher degree of relationship has been reported by other investigators in the case of representative drawing may be explained by the presence of certain more or less common aspects in the measurement of object drawing for accuracy of representation and for the developmental characteristics considered by *Goodenough*.

B. For the artistically superior high school subjects and adult artists.

(1) There is a tendency for artistically superior subjects at

the junior and senior high school level to be somewhat superior in intelligence, although not to the degree suggested by *Terman*. The average IQ for the group is near the dividing line between ' high average ' and ' superior ' in the classification as given by *Terman*. This position on intelligence is not commensurate with the subjects' manifest superiority in artistic ability.

(2) This same tendency is apparent in the adult artist group although more markedly than in the case of the high school subjects, the average IQ for the artists falling near the line separating ' superior ' from ' very superior.'

(3) While the IQ average of the artist group is high, it does not necessarily follow that there is a direct relationship between intellectual status and importance as an artist. The rankings of these artists by two experts, furthermore, suggests that such a relationship is doubtful.

C. General summary. In considering the results of both phases of the study the following summary of conclusions may be made:

(1) Artistic ability is a special ability in the sense of being only somewhat related to general intelligence as measured by established tests. This conclusion, however, applies primarily to the normal group.

(2) There is a tendency for a higher than average degree of intelligence to be present with artistic superiority as indicated by the average of IQ's found in the case of the selected groups. It appears that the use of artistic potentialities in the achievement of marked success in art *is dependent to a certain extent on intellectual capacity.* While no data having direct bearing on this point were obtained in the present study, results from other investigations, particularly in the field of music, lend confirmation to such an interpretation of the function of intelligence in artistic achievement.

Bibliography

1. American Art Annual (Ed. McGlauflin, A. C.). Washington, D. C., American Federation of Arts, 1931.
2. Ayer, F. C. The psychology of drawing. Baltimore, Warwick & York, 1916.
3. Bird, M. H. A study in æsthetics, *Harvard Monog. in Educ.*, 1932, No. 11.

4. Burks, B. S., Jenson, D. W., Terman, L. M., *et al.* Genetic studies of genius, Vol. 3, The promise of youth. Stanford Univ., Calif., Stanford Univ. Press, 1930.
5. Chandler, A. R. Beauty and human nature. N. Y., D. Appleton-Century Co., 1934.
6. Cox, C. M. Genetic studies of genius, Vol. 2. The early mental traits of three hundred geniuses. Stanford Univ., Calif., Stanford Univ. Press, 1926.
7. Dewey, J. Art as experience. N. Y., Minton, Balch & Co., 1934.
8. Elderton, E. On the association of drawing with other aptitudes in school children, *Biometrika*, 7, 1909, 222–226.
9. Fischlovitz, A. An inductive study of the abilities involved in drawing. N. Y., Columbia Univ., 1903.
10. Garrett, H. E. Statistics in psychology and education. N. Y., Longmans Green, 1926.
11. Goodenough, F. L. Measurement of intelligence by drawings. Yonkers-on-Hudson, N. Y., World Book, 1926.
12. Goodenough, F. L. Children's drawings. [In] A Handbook of Child Psychology. Worcester, Mass., Clark Univ. Press, 1931, 480–514.
13. Griffith, C. R. An introduction to applied psychology. N. Y., Macmillan, 1934.
14. Hollingworth, L. S. Special talents and defects. N. Y., Macmillan, 1925.
15. Hollingworth, L. S. Gifted children: their nature and nurture. N. Y., Macmillan, 1926.
16. Ivanoff, E. Recherches experimentales sur le dessin des écoliers de la Suisse romande: Correlation entre l'aptitude au dessin et les autres aptitudes, *Arch. de Psychol.*, 8, 1909, 97–156.
17. Kerchensteiner, I. G. Die Entwickelung der zeichnerischen Begabung. Munich, Gerber, 1905.
18. Kik, C. Die übernormal Zeichenbegabung bei Kindern, *Zeits. für angew. Psychol.*, 2, 1909, 92–149.
19. Kinter, M. The measurement of artistic abilities. N. Y., Psychological Corp., 1933.
20. Kline, L. W., and Carey, G. L. A measuring scale for freehand drawing, *Johns Hopkins Stud. in Educ.*, 5, Part I, 1922.
21. Kline, L. W., and Carey, G. L. A measuring scale for freehand drawing, *Johns Hopkins Stud. in Educ.*, 5, Part II, 1933.
22. Kuhlmann, F. The Kuhlmann-Anderson intelligence tests compared with seven others, *J. Appl. Psychol.*, 12, 1928, 545–594.
23. Kuhlmann, F., and Anderson, R. G. Intelligence tests. Instruction manual (4th ed.), Minneapolis, Educational Test Bureau, 1933.
24. Lewerenz, A. S. I.Q. and ability in art, *School & Soc.*, 27, 1928, 489–490.
25. McCarty, S. A. Children's drawings. Baltimore, Williams & Wilkins, 1924.
26. Manuel, H. T. Talent in drawing. Bloomington, Ill., Pub. Sch. Publ. Co., 1919.
27. Manuel, H. T., and Hughes, L. S. The intelligence and drawing ability of young Mexican children, *J. Appl. Psychol.*, 16, 1932, No. 4, 382–387.
28. Meier, N. C., and Stinson, R. E. Æsthetic sensitivity and productive ability. (In progress; to be published subsequently.)
29. Murphy, G. General psychology. N. Y., Harper & Bros., 1933.
30. National Academy of Science Memoirs. Psychological examining in the United States army, Vol. 15. (Ed. Yerkes, R. M.) Washington, Government Printing Office, 1921.

31. Otis, A. S. Otis self-administering tests of mental ability. Manual of directions and key (for intermediate and higher examinations). Yonkers-on-Hudson, N. Y., World Book, 1928.
32. Pelikan, A. G. The art of the child. Milwaukee, Bruce Publ. Co., 1931.
33. Saunders, A. W. The stability of artistic aptitude at the child level (this volume, pp. 126–154).
34. Skalet, M. A statistical study of the responses of a group of normal children to the individual tests in the Stanford revision of the Binet-Simon scale, *Psychol. Clinic*, 21, 1932, No. 3, 183–195.
35. Stanton, H. M. Measurement of musical talent, *Univ. of Iowa Stud.*, Studies in the Psychology of Music, 2, 1935.
36. Terman, L. M. The measurement of intelligence. Boston, Houghton Mifflin, 1916.
37. Terman, L. M., *et al.* Genetic studies of genius, Vol. 1. Mental and physical traits of a thousand gifted children. Stanford Univ., Calif., Stanford Univ. Press, 1925.
38. Terman, L. M., and Burks, B. S. The gifted child. [In] A Handbook of Child Psychology (2nd ed.), Worcester, Mass., Clark Univ. Press, 1933, 773–801.
39. Tiebout, C. E. The measurement of quality in children's paintings by the scale method (this volume, pp. 85–94).

THE STABILITY OF ARTISTIC APTITUDE AT THE CHILDHOOD LEVEL[1]

by

AULUS WARD SAUNDERS

I. Introduction. In the realm of special ability there is a widespread belief that artistic ability is inborn and remains more or less fixed throughout life. Among many art teachers, on the other hand, the belief is likewise widely held that anyone can learn at least the skills utilized in drawing and painting. The objective of this study was to test the possibility of art ability[2] at the childhood level being changed, with an attempt to identify the factors that might be involved if such a change were brought about. In other terms, the problem was as stated by *Kinter* (27), " Is artistic aptitude relatively fixed and stable, or can it be altered by intensive training? "

In general the prevailing attitude in the field of art psychology has been that an individual is born with a definite capacity for artistic appreciation and production and that this capacity is bounded by limits, some point within which he will succeed in attaining, depending on the nature of his environment and his endeavor. This investigation has been directed toward the nature of these capacity-limits, in an attempt to discover whether it is possible to produce a definite extension of the usual limits. This plan was specifically attempted, first, through the use of a number of subjects who, at the start of the investigation, were definitely inferior, and a few who were definitely superior in art ability, and secondly, through the examination of a section of the general population at the same age-level.

---

[1] An investigation in the research program, Genetic Studies in Artistic Capacity, under the direction of Dr. Norman C. Meier and made possible by financial aid from the Carnegie Corporation of New York through the Carnegie Foundation for the Advancement of Teaching.

[2] Such terms as ability, capacity, aptitude, and the like are used herein as defined in *English's* " A Student's Dictionary of Psychological Terms " (17). Ability, for instance, is there defined as " actual power, whether inherited or acquired, to perform any act. Capacity, while nearly synonymous, usually means undeveloped power, and is largely innate; under training it becomes ability." For definition of such art terms as balance, rhythm, harmony, *etc.*, see " Report of the Committee on Terminology "(61).

The field of this problem was restricted to school children within the age-range of five to nine years. Children of this age have been found by *Tiebout* (57), *Grippen* (22), *Whorley* (62), and others to be old enough to function artistically, and at the same time the responses of subjects of this age seem to be more naïve and less conditioned by learned responses in the field of art.

II. Procedure and subjects. The approach to the problem was made in two ways: (1) through the stimulation and measurement of a picked group of artistic and inartistic children (hereafter called "Group A"), and (2) through the study of an adequate sampling of the general population at the same age-levels (called "Group B").

The plan, as actually followed, was to test the possibility of change in art ability by selecting a number of otherwise normal children who were definitely inferior in art ability, and exposing them to a two-year period of planned art stimulation to test the possibility of a resultant change in the level of art ability and, in case of change, to identify the factors involved. The stimulation was to include instruction, motivation, and techniques of any reasonable nature designed to influence and stimulate the subjects artistically. Measurement of art status of each subject was to be made at the beginning, mid-point, and end of the period of stimulation by the use of objective methods of evaluating the original, graphic products of the subjects produced under controlled conditions. Added to this list of inferior subjects was a small number of artistically superior subjects to be included in the group during the second year of stimulation. These were studied to see the comparative amount of change which might result. The two types of subjects were to constitute Group A.

Comparison of Group A with a section of the general population, Group B, was to be made in which measures of the performance of Group B were to be obtained in various controlled situations. This would produce standards for the measurement of Group A performances. Group B was to be used, furthermore, in an analysis of various factors differentiating the artistic from the inartistic, thus procuring a knowledge of the factors to be used, in turn, in the stimulation of Group A.

Group A consisted of 14 children, who were at the start of this program (Nov., 1932), enrolled in grades Junior Primary to 3 in the University Elementary School in Iowa City. Data describing these subjects are presented in Table 1. Of this group, half were male and half were female. All 14 subjects were normal, bright (with one exception) children from normal homes in or near Iowa City. The average of intelligence of the group was superior, the A.M. of the IQ's being 121. The only child below normal was a boy with an IQ of 88, who was repeating the second grade. The intelligence quotients were obtained from *Stanford-Binet* tests given on dates near the beginning of the program of stimulation.

Of the 14 subjects, 10 were selected because of their demonstrated inferiority in art ability, and will be called ' Z ' subjects. The remaining four were chosen because of their demonstrated superiority in art, and will be known as ' X ' subjects.

TABLE 1. *Sex, intelligence quotient, chronological and mental ages, estimated art status of group A subjects (Nov., 1932)*

Subj.	School Grade	Sex	CA	IQ	MA	Art Status
1	3	M	8–6	115	9–9	Z
2	3	F	8–10	101	8–11	Z
3	3	M	7–4	114	8–4	Z
4	2	F	7–8	116	8–11	X
5	2	M	6–10	133	9–1	Z
6	2	M	6–11	140	9–8	X
7	2	M	8–4	88	7–4	Z
8	2	F	7–7	120	9–1	X
9	2	M	7–7	111	8–5	Z
10	1	F	6–2	149	9–2	X
11	1	F	6–3	122	7–8	Z
12	1	F	5–11	129	7–8	Z
13	1	M	6–2	131	8–1	Z
14	J.P.	F	5–2	123	6–4	Z
		A.M.=	7–1	121	8–7	

TABLE II. *Z and X subjects of group A compared*

Subj.	No.	Sex: M	Sex: F	Av. CA	Av. IQ	Av. MA
Z	10	6	4	7–1	116.70	8–3
X	4	1	3	7–1	131.25	9–3

The method of selection of the subjects was as follows: subjects 3, 6, 8, 9, 10, and 13 were used previously as X and Z subjects by *Rodgers, Grippen, Tiebout* and *Dow*. These six formed part of a larger list of X and Z subjects studied by these investigators, who used a series of ratings by pre-school teachers, research assistants, and art teachers; evaluations by three members of the art education staff; and performance on the *Kline-Carey* Drawing Scale (29), as methods of establishing the artistic inferiority and superiority of these children. A more detailed treatment of the method of selection is given in *Rodgers* (51) and *Grippen* (22). The remaining subjects, 1, 2, 5, 7, 11, 12, and 14 were selected on the recommendation of the art teachers and the art supervisor of the elementary school, in addition to the previous ratings of these subjects made by the division of art psychology. Since this evidence was not considered sufficient for the use of these children as subjects in this investigation, an additional check on all 'candidates' was made by the investigator in the form of a procedure of ranking of paintings made by the candidates. In this procedure, four additional children were used, making a total of 18 candidates considered. Three experienced supervisors of public-school art, each in a different city, were used as judges who were asked to rank the paintings (free expression, on standard-size paper) of the candidates in order of merit, and to grade the paintings individually as to ability shown, considering the age and grade in school of the child. Only those candidates whose paintings, by a close agreement of the judges, qualified them as distinctly superior or distinctly inferior in art ability were retained for use as subjects. By this process two of the candidates were eliminated. As two more were with-drawn from use by the school authorities for other purposes, a total of 14 qualified subjects remained.[3]

III. The program of stimulation. The program of stimulation of Group A covered, roughly, two academic years of time, starting in November, 1932, and extending through May, 1934. The first-year period extended through May, 1933; the second-year period began in November, 1933. The Z subjects of group A, 10 in number, were used during the entire period; the X subjects were used for stimulation only during the latter year. Loss of subjects due to death, illness, or removal from the school system reduced the total number materially during the period. The time of group membership (not including group absences for short periods of time) for the 14 subjects is indicated in Table IV. (Page 143.)

Thus seven Z subjects were stimulated during both year periods, three Zs during the first-year period only, two X subjects through the second-year period, and two Xs during part of the second-year period. For purposes of measurement, all but two subjects (1 and 12) were available throughout the entire period.

Specifically the program of stimulation consisted of a two-year series of meetings of the group with the investigator, in an art-class situation in which the subjects were instructed, reëducated to some extent, and, for the most part, allowed to express their own ideas in various art media under the observation of, and with the help of, occasional criticism of the investigator. Included in the program in addition were: the use of proper media for direct expression, natural conditions of work, effective motivation, a clinical report of each meeting, and home visitations.

The investigator was a trained art teacher, with six years of art-teaching experience in a junior high school. It is probable that he did not bring into the situation a teaching ability of an extreme degree. In fact, the entire program of stimulation was not abnormal in that, in the estimation of the investigator, the methods, materials, and quality of instruction could be duplicated in any efficient school.

---

[3] It is to be understood that the foregoing list of rankings, ratings, and judgments are not to be considered as the establishment of the exact art status of the various subjects. An objective procedure of measurement was made for this purpose, furnishing a closer estimate of the art status of each subject at various periods of time.

The physical conditions were not ideal, as no art-room, with its proper lighting, water-supply and adequate working and storage space was available. During the course of the investigation, four different rooms were assigned, all rather small, which made necessary the division of the group into two sections, each meeting at a different time. The equipment used consisted of a child's upright easel, chair, and table space for each subject. The art materials included an adequate supply of 16 in. x 22 in. white drawing (book) paper, colored crayons, black sketching sticks, soft, colored chalk, modelling clay, tempera paint in 2-oz. bottles, large (½ in.) sable brushes, and glasses for holding water. The majority of the work was done with the soft chalk and tempera paint.

Each of the two sections of the group met with the investigator twice a week for periods of 30 to 45 min. Since these periods were scheduled as far as possible to coincide with the time of meeting of the school's regular art-classes, practically all of the subjects received only the regular time allotment for art-instruction. Hence the subjects, for the most part, were exposed to art influence during two periods per week.

The class routine followed this plan, in general: the subjects came (or were brought) to the experimental room where they seated themselves in a row of chairs. When all were present each subject took a place at the table (if clay was to be used), or stood at his easel to paint or draw. The child always found the easel stocked with several sheets of paper, water, paint, and a brush on the floor or on the chair or table at the side of his easel. The easels were arranged under favorable lighting conditions and so placed that a subject standing at his easel could not easily see what the other subjects were painting. A moderate amount of quiet talking among the subjects was permitted by the investigator, who strove to attain in the group an atmosphere of naturalness. At the end of the period, the child named his product or products, washed out his brush, dumped his glass of water or boxed his sticks of chalk, and was excused from the room.

The method of instruction used was a combination of the so-called ' free expression ' method and the directed training of observation (representation) type. In this method, as followed in this program, the subjects were allowed to paint or draw their own ideas from a given general theme or subject, working largely from memory and imagination. Criticisms by the investigator were given sparingly—usually only upon request. Help usually took the form of quick drawings made before the class (these being covered before the subject started to work), or in verbal descriptions of possibilities for work. Occasionally a simple, diagrammatic sketch was made before the group to show gross relationships and proportion of objects under discussion, and even a jointed cardboard figure, constructed of simple oval pieces pinned together at the joints, was resorted to in studying the human figure. Such sketches and aids to observation always *followed* the needs and requests of the child; they were never introduced primarily as a subject for discussion or observation. The child was repeatedly urged to express his own ideas in his own way, as it was felt that, in addition to the possible production of such values as the development of the child's self-confidence and pride in his own creation, this type of resultant graphic product would obviously afford a significant indication of the status of the subject's art ability. Thus creative expression was considered by the investigator to be more probably indicative of the art ability of the subject. The following of this practice, assisted at times by common aids to drawing and imagination, expressed both graphically and in clay, constituted in the main the technique of instruction.

One important phase of the program of stimulation was the system of motivation used. The motivation can be classified under four heads: (1) judicious praise, (2) visits to art exhibits, (3) exhibition of subject work, (4) possession of art materials at the subject's home.

Since during the first year all the subjects were Zs, the praise used took the form of casual but definite approval to anything the subject did which showed sincere effort. Occasionally disapproval was used in the reverse circumstance. Poor drawing, color, and other evidences of lack of ability were not brought to the subject's attention. It was attempted to make the subject feel that his efforts, even though different from those of others, still possessed merit. The subjects were permitted to think they had some ability; most of them believed that they had been selected for inclusion in this special class for that reason. The praise factor, as striven for in this program, was directed toward a realization on the part of each subject that his own graphic products were meritorious; if necessary, the investigator gave actual help in painting to make such realization possible.[4] This form of help was given only to prevent the occurrence of an unfortunate effect, or to correct one already made; it was given quickly and quietly. The child was led to believe, if possible, that practically all the work was his own. The investigator was forced to make use of this "breach of teaching ethics" only rarely, a practice which seemed amply justified in this experimental situation.

Visits to exhibitions of oil paintings were occasionally made. During the two years, four such visits were made. The subjects were allowed to look at the paintings as they pleased, attention to the few remarks of the investigator relative to the merits of various elements of the pictures being entirely voluntary with the subject. In discussing these paintings, as in the class work, the investigator strove for æsthetic appreciation rather than an understanding of the degree of realism used, knowledge of the 'story' depicted, or factual material such as names and dates.

Subjects were taken in small numbers on various Saturday mornings to a studio, where they saw large easels, oil paints and unfinished pictures about them. In this environment they were allowed to draw and paint, using easels and materials identical to those employed in the regular class period.

---

[4] Such products were eliminated from the list of paintings used for purposes of establishment of artistic status.

The exhibition of paintings made by the subjects was arranged largely during the second year. This motivating factor took the form of an occasional showing of one or more paintings by a subject to the group or to the pupils and teacher of the subject's school room; this factor also included the painting of a special picture by each subject for his parents, framed with an appropriate mat and delivered to the subject's home immediately preceding Christmas; it likewise included a formal school exhibit of framed paintings by the subjects, hung for a week and a half on the walls of two special school rooms.

After every meeting of the group the investigator kept a record in which clinical notes of the proceedings were recorded. Responses of the subjects in the various situations of successful endeavor, failure, and praise were noted, as well as all situations of æsthetic significance. Assistance in this phase was obtained by the occasional use of a stenographer, who recorded verbatim samples of the subjects' conversation, as well as responses to questions and remarks of the investigator.

A final phase of the program of stimulation was the contact kept with the home of each subject over the entire period. At the beginning of the program an easel and full set of art supplies were taken to each subject's home, where acquaintanceship to some degree with the home conditions of the subject was made and where relations of co-operation with the parents were established. An effort was made to obtain for each child some degree of freedom and interest in painting and drawing at home. In some cases the parents were urged not to send the child to his easel merely to keep him out of mischief, or to prevent him from doing something else that he wanted to do at the time. After the first visit a second was made, during which a prepared list of questions was asked about the subject, and observations were noted. A few additional visits were made to each home to replenish art supplies and to collect paintings. The parents were not informed as to what the exact basis of selection of their child had been.

Group B consisted of approximately 100 pupils each from grades 1 through 4 from the Des Moines Schools. Since Group

B was used only during the second year of the program, grades 1 through 4 paralleled the grades of the special X and Z subjects. Each grade presented approximately 100 children, whose distribution of intelligence scores approached normality. Intelligence was measured by the revised form of the *Kuhlmann-Anderson* group intelligence test, given by Tiebout at the start of the Group B program (Nov.–Dec., 1933). Descriptive data concerning Group B are supplied in the *Tiebout* article.[5]

The program for Group B consisted, in brief, of three steps: (1) the subjects were selected as previously described; (2) all subjects were given an æsthetic sensitivity test; (3) each subject made three original paintings from which his art status was estimated.

Immediately following the selection of subjects for Group B on the basis of IQ, the responses of the subjects to the group æsthetic sensitivity test were obtained. This test had been prepared by the investigator for this program; it consisted of a series of various art situations in which each subject's choice was recorded. Since each subject's art status (painting score) was later to be determined, the present æsthetic test provided possibilities for discovering whether there were æsthetic sensibilities differentiating the artistically superior from the artistically inferior subjects. Information concerning the nature of these differentiating sensibilities was held to be of major importance, both as general knowledge in this field, and as specific assistance which it would offer in the stimulation of Group A subjects, and in interpreting the findings therefrom. The æsthetic test likewise provided an opportunity to study the growth of æsthetic sensitivity in average children from year to year, furnishing a standard of performance against which the comparison of the special subjects constituting Group A might be made possible.

The art ability-status of the Group B subjects was obtained from a procedure of objective evaluation (described in the following section) of three free paintings made by each of the subjects. These paintings were made under set conditions. In Grade 1-B (B=first semester) two paintings and two crayon

---

[5] Tiebout, C. E., pp. 87–89. THIS VOLUME.

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

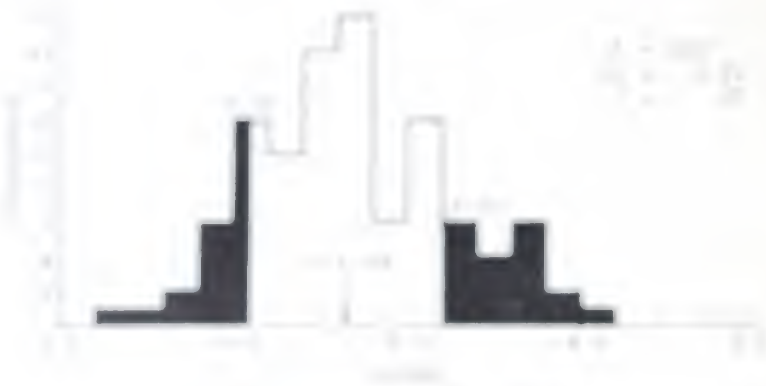
[illegible]

[illegible]

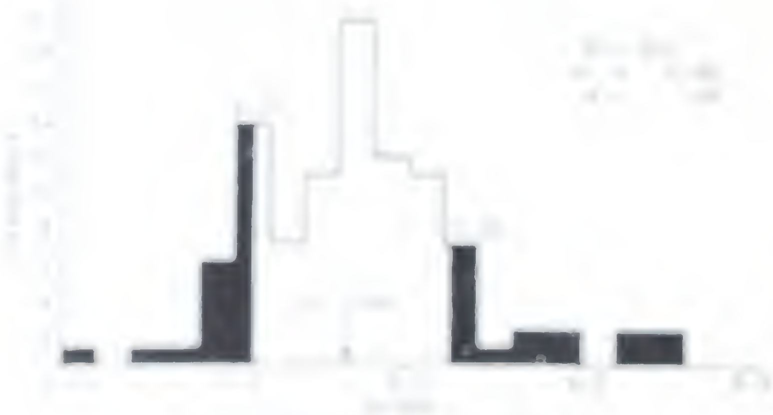
[illegible]

[illegible]

[illegible]



[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

himself, is also obvious. It is believed that in such a manner an inestimable amount of time would be saved, considerably more ground would be covered, and new organizations created that otherwise would have been unknown.

The study of creative imagination in children is now in progress and will be reported in a subsequent volume.

THE LIBRARY

The Ontario Institute

for Studies in Education

Toronto, Canada

American Psychological Periodicals

American Journal of Psychology—Ithaca, N. Y.; Cornell University.
[illegible]

Journal of Genetic Psychology—Worcester, Mass.; Clark University Press.
[illegible]

Psychological Review—Princeton, N. J.; Psychological Review Company.
[illegible]

Psychological Monographs—Princeton, N. J.; Psychological Review Company.
[illegible]

Psychological Index—Princeton, N. J.; Psychological Review Company.
[illegible]

Psychological Bulletin—Princeton, N. J.; Psychological Review Company.
[illegible]

Archives of Psychology—Columbia University P. O., New York City.
[illegible]

Journal of Abnormal and Social Psychology—Eno Hall, Princeton, N. J.
[illegible]

Psychological Clinic—Philadelphia; Psychological Clinic Press.
[illegible]

Journal of Educational Psychology—Baltimore; Warwick & York.
[illegible]

Psychoanalytic Review—Washington, D. C.
[illegible]

Journal of Experimental Psychology—Princeton, N. J.
[illegible]

Journal of Applied Psychology—[illegible]
[illegible]

Journal of Comparative Psychology—Baltimore; Williams & Wilkins Co.
[illegible]

Comparative Psychology Monographs—Baltimore; Johns Hopkins Press.
[illegible]

Genetic Psychology Monographs—Worcester, Mass.; Clark University Press.
[illegible]

Psychological Abstracts—Eno Hall, Princeton, N. J.
[illegible]

Journal of General Psychology—Worcester, Mass.; Clark University Press.
[illegible]

Journal of Social Psychology—Worcester, Mass.; Clark University Press.
[illegible]